ON-LINE PREDICTION OF SOME PROPERTIES OF CRUDE FRACTIONATION PRODUCTS USING ARTIFICIAL NEURAL NETWORKS

A thesis submitted

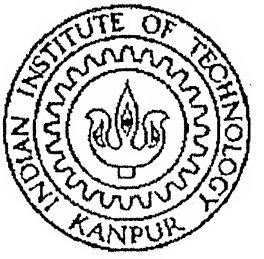
in partial fulfillment of the requirements

for the degree of

MASTER OF TECHNOLOGY

by

TITHIPARNA SENGUPTA



to the

DEPARTMENT OF CHEMICAL ENGINEERING

INDIAN INSTITUTE OF TECHNOLOGY, KANPUR

JULY, 2000

CERTIFICATE

This is to certify that the work contained in the thesis entitled "ON-LINE PREDICTION OF SOME PROPERTIES OF CRUDE FRACTIONATION PRODUCTS USING ARTIFICIAL NEURAL NETWORKS", has been carried out by Ms. Tithiparna Sengupta, under my supervision and has not been submitted elsewhere for a degree.

D.N.Saraf

(Dr. D. N. Saraf)
Professor
Department of Chemical Engineering
Indian Institute of Technology
Kanpur – 208016, INDIA

July, 2000

Acknowledgements

To my thesis supervisor Dr. D. N. Saraf, for the valuable advice and encouragement that I got from him throughout. My gratitude to him is immense and cannot really be expressed in words. I consider it a rare privilege that I could work under him.

To all my lab mates. The camaraderie that we have in our lab was one of the major sources of sustenance in moments of trial. The great time that I had here with my lab mates will be remembered, forever.

To Chauhanji, for his generous help with the hardware and putting back smiles on our panic-stricken faces, so many times.

To Sharmaji, Singhji and Sanjay for their warmth and sincerety.

To all my friends – within and outside IITK.

To IITK itself – for its stimulating academic environment. And also for gifting me with many special moments. It was here that several new friendships, friendships for life, "happened", and old ones were "rediscovered". A special word of thanks to all the trees and birds in the campus, they are very much a part of the happiness here.

To the little children of Jagriti who brighten up life further with their songs and laughter.

To my other two alma maters – South Point School and Jadavpur University. For those lovely memories that can see one through all the rough rides in life.

To my brother, for being the great friend that he is to me.

And finally, to my parents to whom I owe 'all things bright and beautiful' in this world.

Contents

Appendices

List of Tables

List of Figures

Nomenclature

$c_{pl,avg}$ mass average heat capacity of the overflash and the sidestripper liquid products, [Btu/lb-°F]

$c_{pl,i}$ heat capacity of ith side-stripper liquid product(i=1, 2,3 and 4), [Btu/lb-°F]

$c_{pl,m}$ heat capacity of the liquid drawn for mth pumparound (i=1, 2,3 and 4), [Btu/lb-°F]

$c_{pl,ref}$ heat capacity of the liquid reflux to the column, [Btu/lb-°F]

$c_{pv,o}$ heat capacity of the overhead vapor leaving the top of the column, [Btu/lb-°F]

f_c mass flow rate of the feed to the column, [lb/h]

f_{dl} mass flow rate of the liquid distillate product, [lb/h]

f_{dv} mass flow rate of the vapor distillate product, [lb/h]

f_l sum of mass flow rate of the total side-stripper liquid products and overflash that are in vapor form in the flash zone $\left(f_l = \sum_{i=1}^{N_{SS}} f_{SS,i}\right)$, [lb/h]

f_o mass flow rate of the overhead vapor leaving top of the column, [lb/h]

$f_{p,m}$ mass flow rate of the liquid drawn for mth pumparound, [lb/h]

f_{ref} mass flow rate of the liquid reflux to the column, [lb/h]

$f_{ss,i}$ mass flow rate of the ith side-stripper liquid product (i=1, 2, 3 and 4), [lb/h]

F_{dl} molar flow rate of the liquid distillate product, [lbmol/h]

F_L molar flow rate of the total side-stripper products and overflash that are in vapor form in the flash zone, [lbmol/h]

F_o molar flow rate of the overhead vapor leaving top of the column, [lbmol/h]

$F_{ss,i}$ molar flow rate of the ith side-stripper liquid product (i=1,2,3 and 4), [lbmol/h]

F_{ref} molar flow rate of the liquid reflux to the column, [lbmol/h]

H^v_{avg} mass average of the heat of vaporization of the overflash and the side-stripper liquid products, [Btu/lb]

H^v_i heat of vaporization of the ith side-stripper liquid product (i=1,2,3 and 4), [Btu/lb]

H^v_{ref} heat of vaporization of the liquid reflux to the column, [Btu/lb]

L_i total liquid molar flow rate from ith stage, [lbmol/h]

$L_{r,i-1}$ molar flow rate of the hydrocarbon liquid to the ith side stripper draw tray from the tray above, [lbmol/h]

MW_{dl} molecular weight of the liquid distillate

N_{SS} Number of side-stripper liquid products including the overflash

N_{PA} Number of pumparounds

$p_{ss,i}$ partial pressure of hydrocarbon vapor at the ith side stripper draw tray (i=1, 2, 3 and 4), [psia]

p_{top} hydrocarbon partial pressure at the top of the column, [psia]

p_{fz} hydrocarbon partial pressure at flash zone, [psia]

P_{top} total pressure at the top of the column, [psia]

$P_{ss,i}$ total pressure at the draw tray of the ith side-stripper(i=1,2,3 and 4), [psia]

Q_m heat duty of mth pumparound, [Btu/h]

S_i molar flow rate of the steam input to the ith side-stripper (i=1,2,3 and 4), [lbmol/h]

S_R molar flow rate of the steam input to the bottom of the main column, [lbmol/h]

S_T molar flow rate of the total steam input to the column ($S_T = S_T + \sum S_R$), [lbmol/h]

T_{fz} temperature in the flash zone, [K]

T_{ref} return temperature of the liquid reflux to the column, [K]

T_i temperature of the ith side-stripper liquid product (i=1,2,3 and 4), [K]

T_o temperature of the overhead vapor leaving the top of the column, [K]

T_{oc} EFV temperature of overhead vapor (dew point of overhead vapor), [K]

T_{fzc} EFV temperature of the overflash in the flash zone (bubble point of the overflash liquid in flash zone), [K]

$T_{out,m}$ temperature of the liquid stream leaving the column to the mth pumparound, [K]

$T_{in,m}$ temperature of the liquid stream entering the column from the mth pumparound, [K]

V_i total vapor molar flow rate from the ith stage, [lbmol/h]

$V_{ss,i}$ molar flow rate of hydrocarbon vapor and steam leaving the ith sidestripper draw tray, [lbmol/h]

Abstract

The products from the Crude Distillation Unit (CDU) of a petroleum refinery need to conform to certain standards and hence it is important that the properties, which characterize these products, be measured online so that necessary control actions can be taken through feedback mechanism to ensure desired quality. In the absence of availability of online hardware sensors, software sensors need to be developed for the online prediction of product properties. Artificial Neural Networks based models have been developed for the prediction of ASTM temperatures of Heavy Naphtha and the ASTM temperatures, Flash Point and Specific Gravity of Superior Kerosene/Aviation Turbine Fuel. These properties depend not only on the operating conditions in the CDU but also on the type of crude oil that is being processed and its TBP curve, which again is not available and needs to be predicted online. Certain modifications were made on an already existing ANN model, which predicts the feed TBP when supplied with only the operating conditions and the crude type, and several new ANN models were developed that predict the product properties when supplied with the operating conditions and the 'backcalculated' TBP curve predicted by the first model. The properties predictions are generally satisfactory with stray bad results. Since only online operating data are required as the inputs, the package is amply suited for its online implementation. With ANN models working as virtual online analysers, it should be possible to use the information for feedback control and online optimisation of CDU.

Chapter 1

Introduction

Crude distillation is a very important operation in any petroleum refinery. It is the process by which crude oil, which is a mixture of several hundred hydrocarbons of different molecular weights, is separated into different products according to their boiling ranges. The products of crude distillation are, chiefly, Liquified Petroleum Gas (LPG), Straight Run Naphtha (SRN), Heavy Naphtha (HN), Superior Kerosene (SK), Aviation Turbine Fuel (ATF), Light Gas Oil (LGO), Heavy Gas Oil (HGO) and Reduced Crude Oil (RCO). These products are further processed downstream into a host of marketable end products.

Most of the products of crude distillation are characterized by certain properties like RVP for SRN, Flash Point for HN, SK/ATF, LGO, HGO, Pour Point for LGO, HGO and Recovery at 366°C for HGO and RCO at steady state conditions. The ASTM temperatures of the products are also important parameters that characterize them. All these products need to conform to certain standards and this calls for prediction of these properties so that necessary control actions through feedback mechanism may be taken to prevent quality give away. But, unfortunately no online sensors are available for measuring these properties and the off-line laboratory measurements of the properties, being cumbersome and time consuming, cannot be used for feedback control. Hence, the need for software sensors, which will estimate the product properties online, thus enabling effective control and optimization.

The product properties depend on the operating conditions of the Crude Distillation Unit (CDU) as well as on the characteristic of the crude oil being processed. The crude is characterized chiefly by its True Boiling Point (TBP) analysis. The TBP curve is a plot of boiling point temperature vs. volume percent distilled. The product properties are a strong function of the TBP curve of the feed crude oil. However, the TBP data for the pure crudes that are available cannot be used for the purpose of prediction of product properties for several reasons. A typical refinery processes different types of crudes based on economic consideration and availability. In the refinery, crudes are switched frequently, and thus the crude being processed is rarely pure. In the crude storage tanks incoming crudes mix with the residual crude present which maybe of a different type Also, stratification in the storage tanks changes the composition of the crude over the period of time. Moreover, some of the imported

crudes are themselves mixtures of crudes of different types. Taking into consideration all these factors, it is difficult to accept the available original TBP curves for the pure crudes for predicting product properties. This makes it necessary to develop a method to estimate crude TBP online so that this estimated feed TBP curve, along with the operating conditions, can be used for predicting the properties of the products correctly.

An online crude TBP estimation methodology was developed earlier (Murtuza, 1999; Satyadev, 1998) which makes use of measured temperatures at a few stages and other operating conditions. The measured temperatures were corrected for prevailing hydrocarbon partial pressures to obtain EFV (equilibrium flash vaporization) temperatures of different products which were then empirically corrected to crude TBP temperatures using linear functional dependence on a)sum of the mass flow rates of the total side-stripper liquid products and overflash, b)mass flow rate of the overhead vapor leaving top of the column, and c)mass flow rate of the feed to the column. A steady state package called SIMULONREF already exists, consisting of this model for online backcalculation of the feed TBP, and a transport phenomena based steady state model for the CDU (Ramaswamy, 1999 and Roy Choudhury, 1998). However, to do away with the linear approximations of empirical correlations present in SIMULONREF, artificial neural network (ANN) based models are being tried out, for online estimation of feed TBP and subsequent estimation of product properties based on the CDU operating conditions and the ANN predicted feed TBP. While ANN is a totally empirical approach, unlike regressed models it can account for undefined non-linearities with ease and without additional computational burden.

An ANN based model for estimating feed TBP was developed in an earlier work (Das, 2000). The present work involves making certain modifications to this already existing model and also integrating it with some newly developed neural nets so as to develop a package that estimates some of the product properties with only the operating conditions of the CDU, the type of crude being processed, and the original crude TBP being supplied as inputs. The very nature of the inputs for the package ensures that it can be used for online prediction purpose. The ANN toolbox of Matlab (MathWork Inc. USA) has been used for the present work.

An overview of artificial neural net based modeling is presented in Chapter-2.The method of online estimation of feed TBP curve using ANN model is presented in Chapter3.Chapter-4 describes the ANN models that predict the product properties

with the operating conditions of CDU and the ANN estimated feed TBP data as inputs. Finally, conclusions and recommendations for future work are presented in Chapter-5.

Chapter 2

Artificial Neural Network

Artificial Neural Network is a system loosely modeled on the pattern of the human brain. Just as the human brain learns through examples and experience and assimilates the knowledge for future use, artificial neural nets also, through repetitive training by examples, can be "trained" to "learn" the nature of relationship between a set of inputs and outputs and later made to "recall" this relationship to predict the outputs given a new set of inputs. Thus ANN can be defined as a mathematical tool that recognizes a pattern in a data set and builds an internal model of the process generating the data. After initial training, if more data are supplied and the net is retrained, the model will be altered suitably to incorporate the additional learning.

This field of neural computing is a very fast-growing field in the area of Artificial Intelligence and holds out a lot of promise chiefly because of its ability to learn highly complex and non-linear relations with ease. Neural nets can handle very noisy data as well. Moreover, they do not require a prior fundamental understanding of the process. In this regard, it will not be out of context to compare ANN based modeling with regression based modeling. Regression models require the user to specify the functions over which the data sets are to be regressed. In order to specify the functions, the user has to know the form of equations governing the correlations between the data and also needs to have a reasonable numerical and mathematical expertise. Neural nets neither require the specification of the forms of correlations nor any mathematical and numerical expertise.

2.1 Literature review

Research in the field of Artificial Neural Network can be traced back to as far as the 1940-s and a brief summary of it can be found in Das's M.Tech thesis (On-line Crude TBP Estimation Using Artificial Neural Network, 2000). This section strictly deals with the research work done in this field for solving chemical engineering problems. A lot of research has been done and currently underway in the application of ANN in the area of chemical engineering. These are briefly summarized below in chronological order:

- Bhat and McAvoy (1990) discussed the backpropagation modeling approach to model the dynamic response of pH in a CSTR.

- Bhat and Mcavoy (1992) discussed an approach to neural network reduction for modeling.
- Hernandez and Arkun (1992) discussed the use of ANN to learn inverse dynamic models for control.
- Chitra (1993) discussed neural networks for problem solving.
- Brambilia and Trivella (1996) presented an application of neural networks to predict the Octane number of a catalytic reformer stream and product quality of a gas splitter.
- Sadhukhan (1997) used ANN based modeling to predict the properties of petroleum fractions from a crude distillation unit.
- Sharma (1998) used ANN techniques for estimation of phase equilibrium constants of electrolyte systems.
- Shene et al. (1999) discussed ANN based modeling for the prediction of the main state variables in batch fermentations.
- Das (2000) used ANN based modeling for online estimation of crude TBP curve.

The above list is by no means exhaustive and provides only some of the applications for which ANN has been used.

It can thus be seen that despite the interest shown in this field by the chemical engineering fraternity not much work has been done in the area of crude distillation or software sensor development. Sadhukhan's models correlate the properties of the products of CDU with parameters that are measured only off-line and thus they cannot be used for online prediction purposes. The work of Das, involving the development of a methodology for online feed TBP estimation provides the basis for the present work.

2.2 Neural Network Overview

The basic processing element of an artificial neural network is the neuron. Like its biological counterpart it receives and transmits signals, or to put it simply, receives information in the form of data, processes it and transfers the processed data to the next layer. The neurons are clustered into slabs, which form the three layers- the input layer, the hidden layer and the output layer. There may be more than one slab in a particular layer, though normally both the input layer and the output layer consist of single slabs only. The slabs in a particular layer may be arranged in series or in

parallel (Fig: 2.1). The flow of information is in the following order: from the external environment to the input layer, from the input layer to the hidden layer, from the hidden layer to the output layer, and from the output layer to the environment. Each neuron is characterized by what is called its activation function and neurons in the same slab have the same activation function. Each neuron of a particular slab is connected to all the other neurons of slabs in layers previous or next to its own by connections known as "weights" in ANN parlance. The weights are initially randomized and later changed as the "training" of the net proceeds.

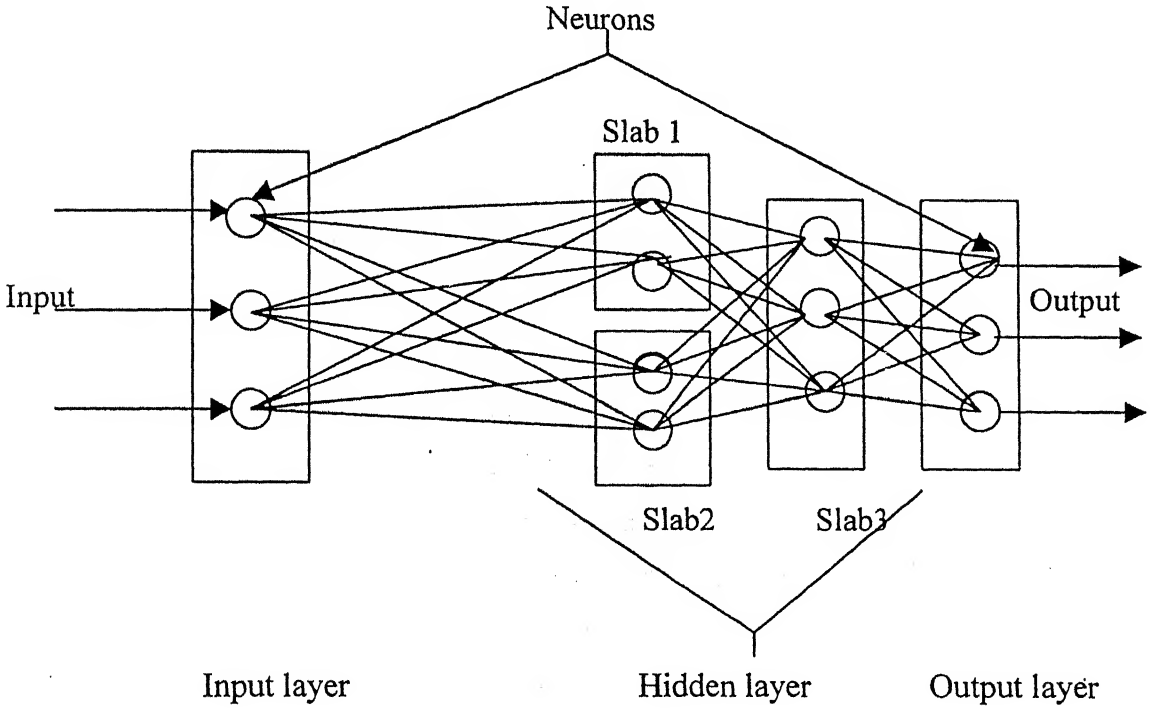


Fig 2.1 Network Structure

The input and the output layers have as many neurons each as the number of input and output variables respectively. Each neuron of the input layer thus receives one input and after multiplying it with suitable weights passes the values to the neurons in the next (hidden) layer. Each neuron in the hidden layer thus receives weighted values from the previous layer, sums these values, and transforms this sum by means of the activation function, before transmitting it to the neurons in the next layer. Thus every neuron in the hidden layer receives several inputs but produces a single output, which it passes on to the different neurons in the output layer after multiplying by different weights. Each neuron in the output layer, in turn, sums all the weighted signals and

after applying the activation function on this sum, produces the output. We have thus as many outputs as the number of neurons in the output layer. The actual desired values of the outputs are also fed to the neural net so that the errors between the results of the output neurons and the desired corresponding "target" values can be computed and propagated backwards through the net to adjust the weights. This is precisely what is the training by backward propagation of error signals to update the connection weights. Repeated forward and backward sweeps through the net, using different input and output data sets, result in a converged set of the weights, yielding a net that is trained to identify patterns between sets of input data and corresponding sets of target values.

2.3 Backpropagation Algorithm

The mathematical algorithm for the backpropagation process is defined by the learning rule used to update the weights. There are several major learning rules like the Hebb's rule, the Hopfield rule and the Delta rule, out of which the last is the most commonly used one. This rule is based on the idea of continuously modifying the strengths of the input connections to reduce the difference (the delta) between the desired output value and the actual output of a neuron. This rule changes the connection weights in the way that minimizes the mean squared error of the network. The error is back propagated into previous layers, one layer at a time. The process of back-propagating the network errors continues until the first layer is reached. The names "Feed forward" and "Back-propagation" are derived from this method of computing the error term. This rule is also referred to as the Windrow-Hoff Learning Rule and the Least Mean Square Learning Rule.

The neural net is trained on sets of data containing input and corresponding target values. The input data could be viewed as an n×p matrix, where n is the number of input neurons and p is the number of different data sets, which are called patterns. A set of m target output values are associated with each of the p patterns, corresponding to m output neurons in the net. In the predictive mode, the net will be presented with one or more patterns of n inputs and will be required to predict m outputs for each input pattern.

The different steps of training the net are as follows:

Step-1: Weight initialization: All weights are initialized to small random numbers.

Step-2: Calculation of activation level: the activation level of an input unit is determined from the input presented to the unit. The activation level $O_{i,p}$ of a neuron i of the hidden or the output layer, for a pattern p, is determined by:

$$O_{i,p} = F(I_{i,p}) \tag{2.1}$$

where $I_{i,p}$ is the input to neuron i for pattern p and is given by:

$$I_{i,p} = \sum_{j}(w_{ji}O_{j,p}) + w_{Bi}O_B \tag{2.2}$$

where the summation over j represents all the neurons in the preceding layer and B represents the bias neuron. The output O_B from this bias neuron is invariant. The weight w_{ji} is associated with the connection from the j^{th} neuron to the i^{th} neuron and w_{Bi} represents the weight of the connection from the bias neuron of the preceding layer to the i^{th} neuron. The output from the neuron in a hidden layer is obtained by substituting the input value obtained from Eq. 2.2 in Eq.2.1. F is the activation function used to transform the sum of the weighted inputs to a neuron.

Step-3: Weight training:

(a) Weight changes are started at the output units and worked backwards to the hidden layers recursively. The weights are adjusted according to the formula:

$$W_{ji}(n+1) = W_{ji}(n) + \Delta W_{ji}(n) \tag{2.3}$$

where $W_{ji}(n)$ was the weight at the start of the ith iteration and $\Delta W_{ji}(n)$ is the weight change in the ith iteration.

(b)Weight change is computed by:

$$\Delta W_{ji}(n) = (\eta\delta_{i,p}O_{j,p}) + \alpha\Delta W_{ji}(n-1) \tag{2.4}$$

where η is a trial independent learning rate ($0<\eta<1$) and $\delta_{i,p}$ is the error gradient at unit i. α is the momentum factor and $\Delta W_{ji}(n-1)$ is the weight change at the $(n-1)^{th}$ i.e. previous iteration.

(c) The error gradient for any neuron i in the output layer is given by:

$$\delta_{i,p} = (T_{i,p} - O_{i,p})F'(I_{i,p}) \tag{2.5}$$

where $T_{i,p}$ is the desired target output for neuron i and pattern p, and F' is the derivative of the activation function used for neuron i.

The error gradient for any neuron in the hidden layer is given by:

$$\delta_{i,p} = F'(I_{i,p})\sum_{k}(\delta_{k,p}w_{ik}) \qquad (2.6)$$

where k represents the neurons to which neuron I in the hidden layer sends its output.

Step-4: Repeat steps 2 and 3 with p incremented by one until all patterns in the training set are exhausted. Several such passes may be required until convergence in terms of error criterion is achieved. An iteration includes presenting a pattern, calculating activations and modifying the weights. When all the patterns in the training set have been iterated upon, it has completed one pass.

2.4 Scaling and Activation Functions

Scaling Functions: When variables are loaded into a neural network, they must be scaled from their numeric range into the numeric range that the neural network deals with efficiently. There are two main numeric ranges the networks commonly operate in, depending upon the type of activation functions used: [0, 1] and [-1, 1].

In addition to the linear scaling functions, there are two non-linear scaling functions: logistic and tanh. The logistic function scales data to (0, 1) according to the following formula:

f(value)=1/(1+exp(-(value-mean)/sd))

where mean is the average of all of the values of that variable in the pattern file, and sd is the standard deviation of those values.

The function Tanh scales to (-1, 1) according to:

f(value)=tanh((value-mean)/sd))

where tanh is the hyperbolic tangent.

Both of these functions will tend to squeeze together data at the low and high ends of the original data range. These nonlinear scaling functions may thus be helpful in reducing the effect of "outliers." They have an additional advantage in that no new data no matter how large is ever clipped or scaled out of range. However, linear scaling is often sufficient and used as default option in MATLAB because of its simplicity.

Activation functions: The neuron performs non-linear operation on its input through the activation function. The activation function, also known as the "squashing function", maps the input into the output value, which is "fired" onto the next layer. Usually it is a sigmoidal function, which is a monotonic, continuously differentiable, bounded function e.g. the logistic function $f(x)=1/[1+\exp(-x)]$. Other activation functions used are tanh, gaussian, gaussian complement etc.(More details about the activation functions available in the MATLAB ANN toolbox are given in Appendix-1)

2.5 Learning Rate and Momentum Factor:

Learning Rate: Each time a pattern is presented to the network, after learning the weights are modified in the direction required to produce a smaller error the next time the same pattern is presented. The amount of modification is learning rate times the error. The smaller is the learning rate η, the smaller will be the changes in the synaptic weights from one iteration to the next and smoother will be the trajectory in weight space. This improvement, however, is attained at the cost of slower rate of learning. On the other hand, if the learning rate parameter is too large so as to speed up the rate of learning, the resulting large changes in the synaptic weights assume such a form that they may become unstable.

Momentum Factor: The momentum factor α provides a smoothing effect to the weight changes and allows the use of larger learning rates. It determines the proportion of the weight change in the last iteration that is added into the new weight change. This variable is needed because large learning rates often lead to oscillation of weight changes and learning never completes, or the model converges to a solution that is not optimum. The value of momentum factor varies from 0.1 to 0.9. Typically, for noisy data, a learning rate of 0.05 and a momentum factor of 0.5 give good convergence.

2.6 Cross-Validation and Overtraining:

A problem that can be faced during training a neural net is overtraining. This results in the net memorizing the training patterns to such an extent that it may fail to predict for other patterns. One approach to avoid over-training of the network is to estimate the generalization ability during training and stop when it begins to decrease.

The essence of back-propagation learning is to encode an input–output relation, represented by a set of data, with a multilayer perceptron (MPL) well trained in the sense that it learns enough about the past to generalize to the future. The simplest method is to randomly partition the data set into a training set and a test set. From the training set a validation subset, which are typically 10 to 20 percent of the training set is set aside. The motivation here is to validate the model on a data set different from the one used for parameter estimation. The training set is used to modify the weights, the validation set is used to estimate the generalization ability, and training is stopped when the error on the validation set begins to rise. Another way of avoiding over-training is to limit the ability of the network to take advantage of spurious correlation in the data. Overfitting is thought to happen when the network has more degrees of freedom (the number of weights, roughly) than the number of the training samples – when there are not enough examples to constrain the network. Even though it may give exactly right output at the training points, it may be very inaccurate at other points. An example is a higher order polynomial fitted through a small number of points.

Sufficient Training Set Size For a Valid Generalization:

Generalization is influenced by three factors: the size and the efficiency of the training set, the architecture of the network, and the physical complexity of the problem at hand. Clearly we have no control over the latter .If the architecture of a network is fixed then the size of training can be derived as follows.

Let M denote the total number of hidden layer computation nodes. Let W and N be the total number of synaptic weights and the number of random examples used to train the network respectfully. Let ε denote the fraction of error permitted on test. Then, according to Baum and Haussler, (Simon Haykins, 1997) the network will almost certainly provide generalization provided the following two conditions are met.

(a) The fraction of error made on the training set is less than $\varepsilon/2$.

(b) The number of examples used in the training is

$$N \geq \frac{32W}{\varepsilon} \ln\left(\frac{32W}{\varepsilon}\right)$$

Ignoring the logarithmic factor, taking first order approximation, the number of training examples is directly proportional to the number of weights in the network and inversely proportional to the accuracy parameter η. Then,

$$N > \frac{W}{\varepsilon}.$$

Chapter 3

Online crude TBP backcalculation

The importance of online estimation of the feed TBP curve for predicting the properties of the products was discussed in Chapter 1. A semi-empirical model for online estimation of feed TBP curve has already been developed in earlier works (Murtuza, 1999; Satyadev, 1998). But considering the convenience of integrating an ANN based model for TBP backcalculation with the ANN based model for property prediction, the problem of TBP backcalculation was addressed again, using ANN. In this context, it will be prudent to discuss the existing semi-empirical model before going into the details of the ANN based model.

3.1 Existing semi-empirical model for TBP backcalculation

This model for on-line crude TBP estimation was developed using a heat balance around the crude column's rectification section. The heat balance calculations give an estimate of the amount of vapor leaving the flash zone. When crude composition changes the amount of vapor increases or decreases, depending on whether the incoming crude is lighter or heavier. When lighter crude enters the column, the amount of vapor leaving the flash zone increases when operated under the same conditions. Thus the tower cooling load, that is the heat duties of the pump-around and the reflux condenser increase proportionally. The increase in the cooling load is independent of the total distillates (products from the column) drawn. If the side stream flows are not increased, the extra amount of vapor, which is condensed in the condenser, is returned to the column increasing the internal reflux on each tray, thereby increasing over-flash.

The above development was based on the work of Friedman (1985) who backcalculated two points on the TBP temperature versus volume percent distilled curve, thus approximating the curve with a straight line. The procedure developed by Satyadev (1998) provides for the estimation of a set of six points, which allows drawing a more realistic TBP curve than a straight line. Figure 3.1 shows the rectifying section of a CDU column. The dotted line shows the envelope around which heat balance is performed. At steady state the total heat balance can be written as:

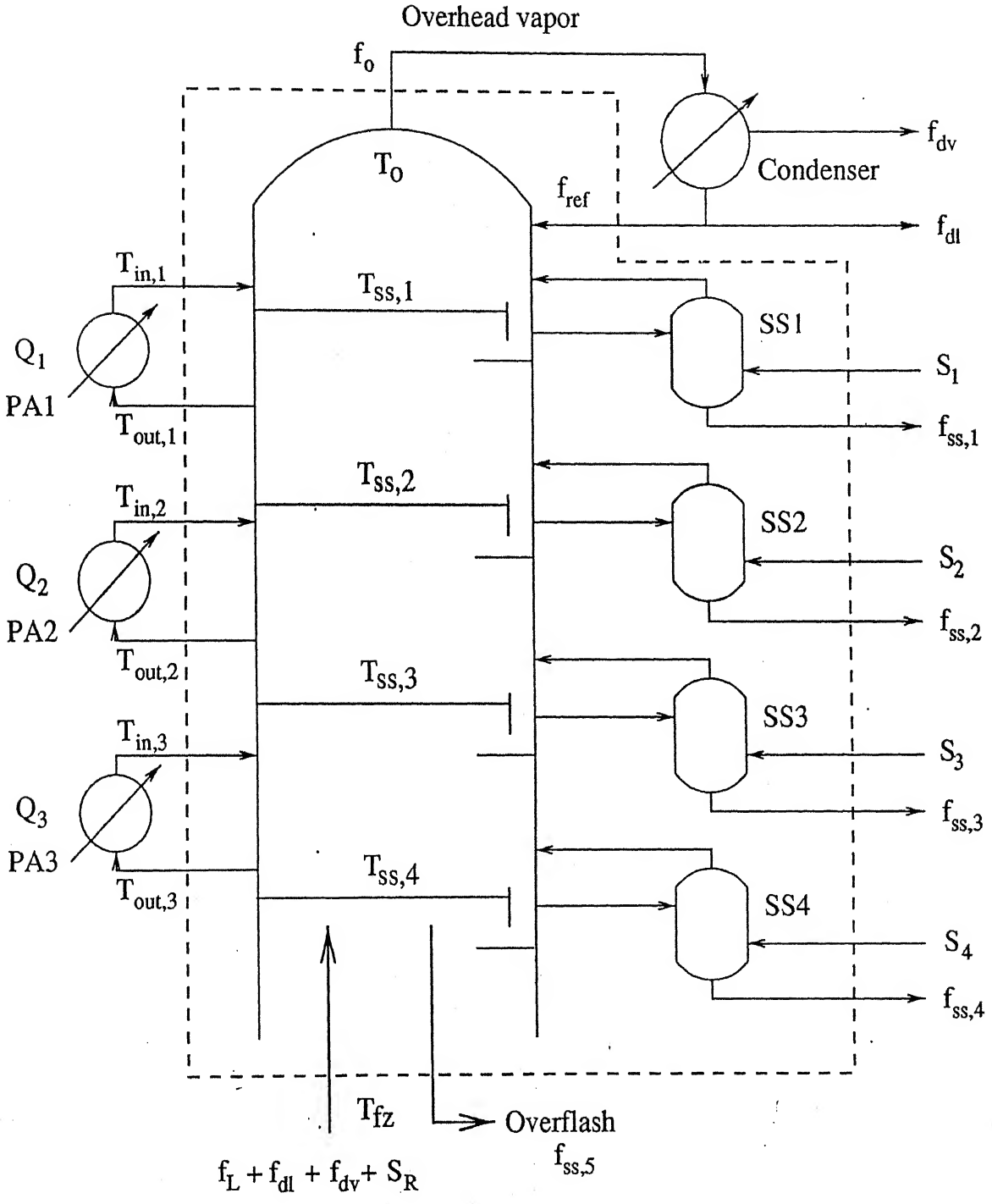


Figure 3.1: Heat balance envelope encompassing the rectifying section of CDU

$$H_{in}-H_{out}=0 \tag{3.1}$$

The enthalpy of the saturated hydrocarbon vapor at the flash zone temperature and pressure defines the chosen reference condition for the enthalpy. This makes the enthalpy entering the envelope with the vapors from flash zone as zero. Also it is assumed that the enthalpy of the stripping steam at the inlet and outlet conditions are roughly the same, hence not considered in the energy balance calculations. The only stream that contributes to the input of enthalpy to the envelope is the reflux stream, which is given below:

$$H_{in}= - [H^{v}_{ref} + c_{pl,ref} (T_{fz} - T_{ref})] f_{ref} \tag{3.2}$$

The enthalpy leaving the column is the sum of the side stream enthalpies, overflash enthalpy, overhead vapor enthalpy and heat duties of the pumparounds.

$$H_{out} = \sum_{i=1}^{Nss} -[H_i^{v} + c_{pl,i}(T_{fz} - T_i)]f_{ss,i} - c_{pv,o}(T_{fz} - T_o)f_o + \sum_{m=1}^{Npa} Q_m \tag{3.3}$$

The pumparound heat duty is calculated from the following expression:

$$Q_m=c_{pl,m}(T_{out,m} - T_{in,m})f_{p,m} \tag{3.4}$$

Assuming that the combined liquid product enthalpy can be calculated as if all the liquid products leave the tower at an average temperature, T_{avg}.

$$\sum_{i=1}^{Nss}[H_i^{v} + c_{pl,i}(T_{fz} - T_i)]f_{ss,i} = H^{v}_{avg} + c_{pl,avg} (T_{fz} - T_{avg}) f_L \tag{3.5}$$

where f_L can be defined as the sum of the total side-stripper products and the overflash stream. The mass average property, for example (T_{avg}) is given as:

$$T_{avg} = \frac{\sum_{i=1}^{Nss} f_{ss,i} T_i}{\sum_{i=1}^{Nss} f_{ss,i}} \qquad (3.6)$$

The overflash liquid is assumed to be at flash zone temperature, implying that on the left hand side of the equation (3.5) $T_i = T_{fz}$, reducing the liquid enthalpy term to zero. The partial pressure of the hydrocarbon product vapors is calculated from the actual tower pressure and the molar flow rate of hydrocarbons are calculated from mass flow rate of the product.

The measured temperatures a) top tray temperature, T_o, b) flash zone temperature, T_{fz}, and c) Side-stripper column draw temperatures, $T_{ss,i}$ (i=1,2,3 and 4) correspond to the hydrocarbon product partial pressures in the column at the appropriate locations. However, EFV and TBP curves are at atmospheric pressure, hence these temperatures have to be corrected for pressure difference.

The partial pressure of the hydrocarbon product vapors is calculated from the actual tower pressure and the molar flow rate of hydrocarbons. The molar flow of these hydrocarbon streams is given by:

molar flow rate = (mass flow rate) / (molecular weight)

For example the molar flow rate of liquid distillate is given by:

$$F_{dl} = f_{dl} / MW_{dl} \qquad (3.7)$$

The partial pressure of the hydrocarbon vapors in the overhead is given by:

$$p_{top} = P_{top}\left(\frac{F_{ref}}{F_O + S_T}\right) \qquad (3.8)$$

The partial pressure of the hydrocarbon product vapors at the flash zone is given by:

$$p_{fz} = P_{fz}\left(\frac{F_L + F_{dl}}{F_L + F_{dl} + F_{dv} + S_R}\right) \qquad (3.9)$$

The partial pressure of the hydrocarbon product vapors above the draw tray for the ith side stripper is given by:

$$p_{ss,i} = P_{ss,i}\left(\frac{L_{r,i-1}}{L_{r,i-1} + V_{ss,i} - F_{ss,i-1}}\right) \qquad (3.10)$$

where $L_{r,i-1}$ is the hydrocarbon liquid reflux flow to the ith side stripper column draw tray (liquid flow from the tray above the draw tray) and $V_{ss,i}$ is the total hydracarbon vapors and steam flow leaving the ith side stripper column draw tray (see figure 3.2). The flow rate of the previous side stripper product ($F_{ss,i-1}$) is neglected from the total hydrocarbon vapor leaving the draw tray. The reason behind this is that the product vapor, $F_{ss,i-1}$ is near its critical temperature as it leaves the draw tray under consideration and hence, it is assumed to have no effect on partial pressure. But the product vapors which are to be withdrawn from the side-strippers above the previous side-stripper, i.e. $F_{ss,i-1}$, are above their critical temperatures when they are at the temperature of this draw tray. Thus, like the steam, they behave as non-condensibles and lower the boiling point of the product in accordance with Dalton's law of partial pressures. For the calculation of hydrocarbon partial pressure at the first side-stripper draw tray, liquid distillate (corresponds to $F_{ss,i-1}$ for this case) must be subtracted from the total hydrocarbon vapors leaving the draw tray. The vapor distillate still behaves as non-condensible and must be accounted for. For the calculation of hydrocarbon partial pressure at the top tray, the expression used here is different from that in the book "Petroleum Refinery Distillation" by R.N. Watkins, where it is suggested that the term for the molar flow rate of the drawn liquid distillate i.e. F_{dl} be also included in the numerator. However, at the temperature and pressure conditions prevailing at the top tray, only the reflux is in the liquid state and the distillate, being in the vapor state can be considered as a non-condensible at that condition. Hence, while calculating the partial pressure the term for the distillate i.e. F_{dl} has been dropped out from the numerator.

The measured temperatures at the partial pressure of the hydrocarbon streams are corrected for the atmospheric pressure using the Clasius-Clayperon equation. Thus the top tray temperature, T_o, when corrected corresponds to the dew point of the overhead vapors (liquid distillate), T_{oc}, while the draw tray temperature $T_{ss,i}$ (i=1,2,3 and 4), corrected, corresponds to the bubble point of the unstripped liquid hydrocarbon on the respective draw trays, $T_{ssc,i}$ (i=1, 2, 3 and 4). The flash zone temperature, T_{fz}, when

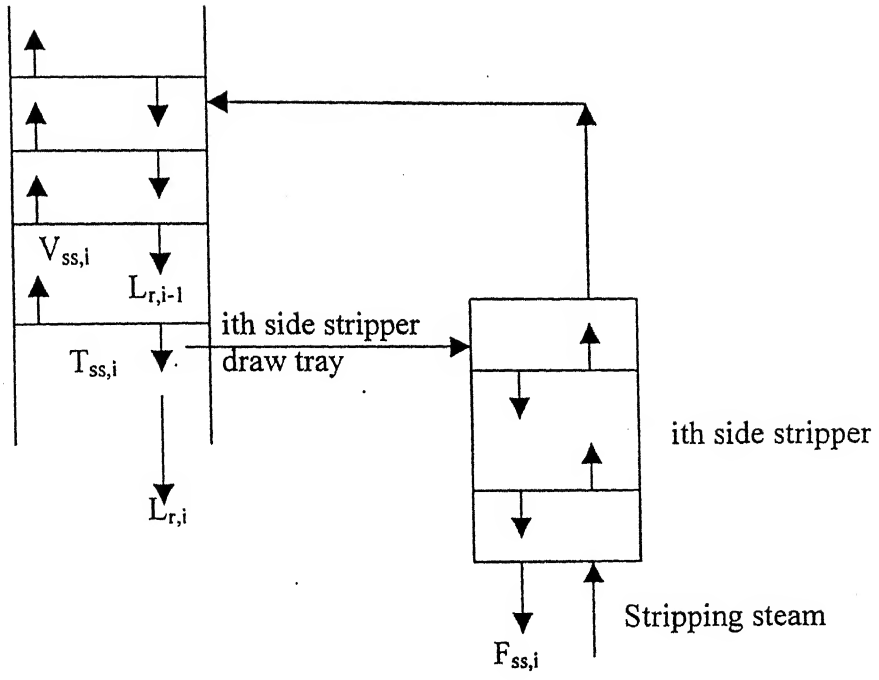


Figure 3.2: Section of distillation column where side-stripper is connected with the main column through various flows

corrected, corresponds to the bubble point of the overflash liquid, T_{fzc}, in the flash zone. Thus these temperatures from the cut points on the EFV curve correspond to the respective cumulative volume percent distilled.

These six EFV temperatures were converted into crude TBP temperatures using an empirical correlation, which was a function of column parameters. The correlation used a linear relationship between the EFV and TBP temperatures with correlation coefficients themselves being linearly dependent on mass flow rate of feed, overhead vapor leaving the top of the column and the total side-stripper liquid products including overflash. Thus six TBP temperatures are obtained corresponding to locations in the distillation column where amounts of distilled hydrocarbons are measured or calculated. As in this method it is not possible to predict the initial and final points on TBP curve, these were taken to be the same as in the initially available laboratory TBP curve. The initial boiling point and the final boiling point as well as the TBP temperature corresponding to 80 volume % distilled, all taken from the original laboratory TBP curve corresponding to the crude concerned, along with the

six TBP points generated by the model, are joined by straight lines so as to generate the complete TBP curve. Actually the span of the gap between the sixth point generated by the model and the FBP is quite large and joining these two points by a straight line is a very crude approximation. So, the 80 volume % point is taken from the original TBP curve, so that it can 'guide' the backcalculated TBP curve to a better shape. However, it is to be understood that the points on the TBP curve beyond the sixth point correspond to the part of the crude that goes into the RCO (Reduced Crude Oil) stream, which is not of any theoretical interest.

The logic behind discussing this semi-empirical model is to get a better understanding of the ANN based model for online TBP estimation, since the latter uses this semi-empirical model to generate data for training the neural net.

3.2 Formulation of the problem

The aim is to develop an ANN based model for backcalculation of TBP curve using the initial TBP curve and the operating conditions as inputs. This new model also finds out the six temperatures at the six locations of the CDU column, but using Neural Network. A model was developed by Das (2000) for this purpose and the present work involves making certain modifications in that model. The previous model as well as the present one is discussed in the following subsections.

3.2.1 Previous model:

Input To The Model: There are 31 neurons in the input layer receiving 31 inputs a) six temperatures which have been taken from the initial TBP curve of the feed crude, corresponding to the volume percent distilled at the six points of the distillation column (the top tray, the four side-streams draw trays and the flash zone) b) feed flow rate, c) flash zone temperature, d) reflux flow rate, e) reflux temperature, f) bottom steam flow rate, g) flash zone pressure, h) three pump-around (HN, KERO, LGO) flow rates and three pump-around return temperatures, i) four side-stripper (HN, KERO, LGO, HGO) product flow rates and four side-stripper steam flow rates, j) top temperature, k) heavy naphtha draw temperature, l) KERO draw temperature, m) LGO draw temperature, n) HGO draw temperature.

Output From the Model: There are six neurons in the output layer to predict six temperatures corresponding to the volume percentages distilled at six points of the distillation column.

Training Set: For training the network, both the inputs and the corresponding outputs must be known. However, since the TBP curve, which is the output of the ANN model, cannot be measured on-line, it must be estimated. This can be accomplished using the backcalculation procedure of Murtuza (1999), already discussed in length in the previous section. However, instead, a reconciliation procedure was used, the details of which can be found elsewhere (Basak, 1998).In this TBP reconciliation procedure, laboratory determined ASTM temperatures of the CDU products are used to reconcile the feed TBP. Since ASTM temperature measurement is an off-line procedure, the reconciliation can not be done on-line. But, for generation of outputs for the training sets, it is a satisfactory method. For the operating conditions corresponding to each data set used for "training" the neural net, the original laboratory TBP curve for the crude concerned was "reconciled" using the program for TBP reconciliation. The temperatures corresponding to the volume percent vaporized at the six points of the distillation column were read off from the "reconciled" TBP curve and used as outputs in the training mode.

Network Architecture:

Supervised type of backpropagation network was used for the model. A network with three slabs in the hidden layer was chosen for this problem - the first slab connected in series with the next two slabs in parallel. Figure 3.3 shows the architecture used. In MATLAB input layer comes by default and so there is no need to assign it separately. Hence, the input layer is not shown in Fig 3.3 or any of the subsequent figures. The different activation functions used are explained in Appendix 1.

The network structure was like this:

Input layer: Number of neurons: - 31.

First hidden layer: Number of neurons: -13

Activation function: 'logsig'.

Second hidden layer:

Slab-1: Number of neurons: -12

Activation function: 'tansig'.

Slab-2: Number of neurons: - 9

Activation function: 'radbas'

Output layer: Number neurons: - 6

Activation function: 'purelin'

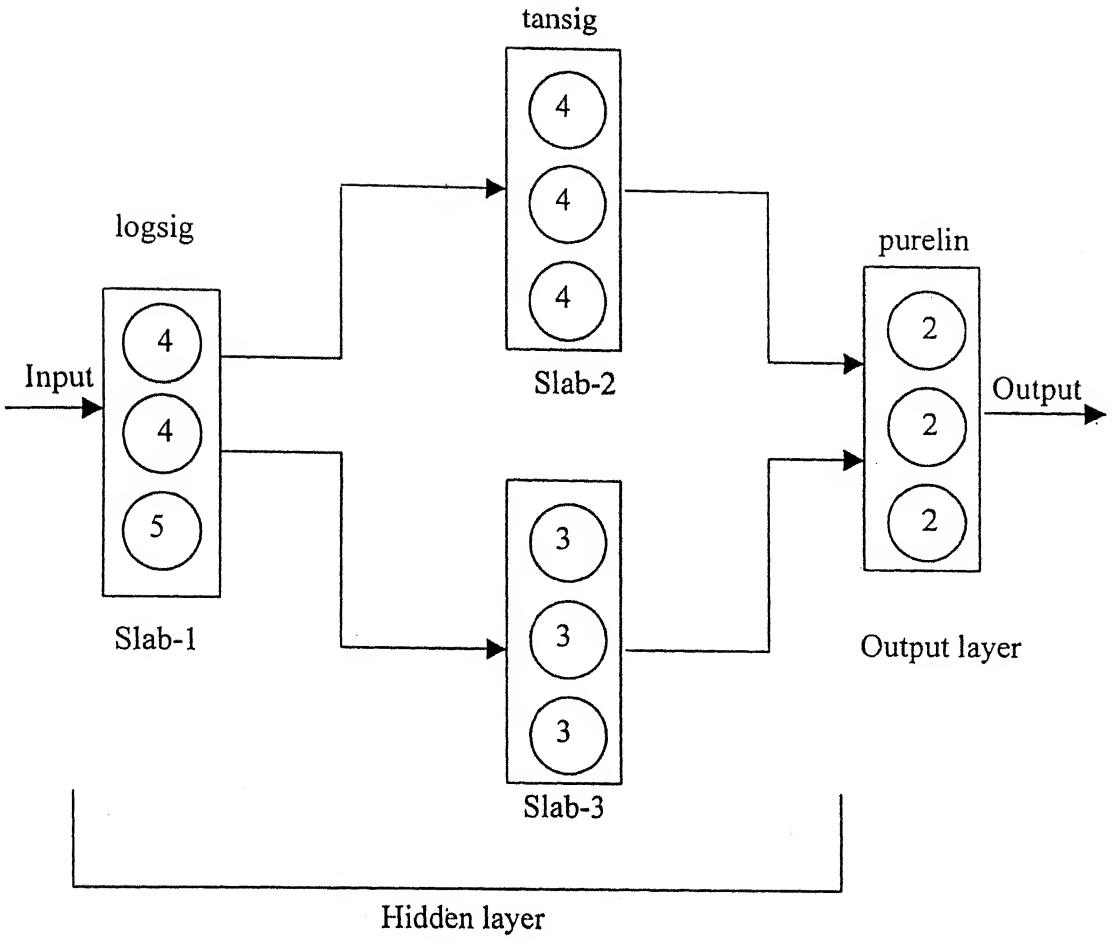


Fig 3.3 Network for predicting crude TBP curve (old model)

'Tansig' function in the second hidden layer was selected to trap the increasing nature of all temperatures with increase in volume percent distilled. It gives faster convergence as compared to 'logsig' during training because of its higher slope. This function also provides larger weightage to input TBP temperatures, because the nature of the initial TBP curve almost matches with the nature of 'tansig' function. The training function used was 'Trainlm' and the performance function was 'mse'. 'Trainlm' is an optimization technique (Levenberg-Marquardt method), which adjusts the weights of the neurons to minimize discrepancy between the target and the prediction. 'mse' is the 'mean square error performance function which really forms the objective function for the Levenberg-Marquardt technique.

However, in the absence of availability of complete ASTM data required for the effective reconciliation of the laboratory TBP curves, there was an inherent error in the TBP reconciliation itself, which resulted in the use of erroneous output data for

training. Hence, it was decided not to use the reconciled TBP curve as outputs for training. This called for making modifications in this ANN model. Hence, the present model was developed.

3.2.2 Present model

The main difference between the present model and the previous one is in the output data used for training. The outputs (target TBP temperatures) were calculated using online backcalculation procedure of Murtuza(1999) as discussed in Section 3.1. This was necessitated in the present work since sufficient laboratory data on product ASTM temperatures were not available for the generation of the reconciled crude TBP curves. The semi-empirical method used here is likely to produce only approximate values and should be used only when laboratory measures ASTM temperatures are not available. Also, in the present model the outputs are five TBP temperatures instead of the six in the previous model. These feed TBP temperatures correspond to the volume percent of the feed crude vaporized at five locations in the distillation column: the top tray and the four side-stream draw trays. The sixth point i.e. the temperature corresponding to the volume percent vaporized in the flash zone was dropped out because of the uncertainty in the backcalculation procedure about correctly calculating this TBP temperature.

The inputs for the neural net in the new model, apart from the operating conditions, are the five temperatures corresponding to the volume percent vaporized at the five locations of the column, read off from the initial laboratory TBP curve of the crude being processed. Some of the other inputs i.e. the operating conditions were clubbed in groups of two or three so as to decrease the number of input parameters. The operating conditions include various stream flow rates and temperatures. Intuitively one would expect the temperatures to be dependent on stream enthalpies rather than mass flow rates. For example one could use input crude enthalpy as a single input in lieu of feed flow rate and its temperature. No attempt was made to calculate the actual enthalpy but the product of feed flow rate and the Coil Outlet Temperature (COT) was taken to be a measure of it. Similarly, the pumparound temperature and flow rate were replaced by a single quantity- [flow rates × (draw temperature – return temperature)]. Similar other groupings resulted in the decrease of input variables from the original 31 to 22. Applying Principal Component Analysis further decreased the number of

input variables. The concepts of normalization and principal component analysis are discussed in the later part of this subsection.

Input to the ANN model: The input variables for this new ANN model are: a)five initial TBP points for five volume percentages b) crude specific gravity c)feed flow rate × COT d)reflux flow × reflux temperature e) HN flow rate × HN draw temperature f) Kero flow rate × Kero draw Temperature g) LGO flow rate × LGO draw temperature h) HGO flow.rate × HGO draw temperature i) three (HN, LGO, HGO) pumparound flow rates × (draw temperature – return temperature) j) four side-stripper steam flow rates k) bottom steam flow rate l) flash zone pressure m)top temperature . Thus there are 22 input variables. The variables along with their ranges of variation are listed in Table3.1. However, after applying principal component analysis the number of variables decreased to 18 and thus the input layer has 18 neurons to receive the 18 inputs.

Output from the ANN model: There are five neurons in the output layer to predict the five temperatures corresponding to the five volume percentages.

Training Set: The different inputs and the outputs for the training set have already been discussed. The selection of training patterns is an important consideration, because the success of the model's prediction capability is dependent on the proper training of the net. Here the training set consists of 51 different patterns taken randomly from the available data set so that it covers the entire range of values and is thus representative of the entire data set.

Validation set: Section 2.6 in the previous chapter underlines the importance of cross-validation to prevent "over-training" of the net. In this case, 19 patterns taken randomly over the entire data set were grouped to form the validation set.

Test Set: The real test of the efficiency of a trained net is in its ability to predict outputs when presented with a set of inputs that it has never seen before i.e. in its ability to generalize. Here, 17 patterns taken randomly over the entire data set form the test set. The net's prediction ability is checked on the basis of its performance on this test set.

Table 3.1: Inputs for prediction of crude TBP temperatures and range of data used

Model Inputs	Input data range for
*Feed TBP temp.(K) corresponding to % vaporization at top tray	389.6-419.5
*Feed TBP temp.(K) corresponding to % vaporization at HN draw tray	420.1-449.1
*Feed TBP temp.(K) corresponding to % vaporization at SK/ATF draw tray	486.1-531.7
*Feed TBP temp.(K) corresponding to % vaporization at HGO draw tray	560.4-603.3
*Feed TBP temp.(K) corresponding to % vaporization at LGO draw tray	588.6-632.4
Crude Specific Gravity	0.824-0.8712
Feed flow rate(m3/h) × COT (°C)	3.97×10^{5}-4.46×10^{5}
Reflux flow rate (m3/h) × Reflux temp. (°C)	5.57×10^{3}-1.38×10^{4}
HN flow rate (m3/h) × draw tray temp.(°C)	8.70×10^{3}-1.40×10^{4}
SK/ATF flow rate (m3/h) × draw tray temp.(°C)	2.34×10^{4}-4.92×10^{4}
LGO flow rate (m3/h) × draw tray temp.(°C)	4.79×10^{4}-6.27×10^{4}
HGO flow rate (m3/h) × draw tray temp.(°C)	1.10×10^{4}-3.23×10^{4}
HN pumparound flow × (pumparound draw tray temp.(°C) – return tray temp. (°C))	1.68×10^{4}-4.14×10^{4}
SK/ATF pumparound flow × (pumparound draw tray temp.(°C) – return tray temp. (°C))	3.73×10^{4}-5.15×10^{4}
LGO pumparound flow × (pumparound draw tray temp.(°C) – return tray temp. (°C))	3.31×10^{4}-5.10×10^{4}
Bottom steam flow rate (tons/h)	6.00-9.96
HN side-stripper steam (tons/h)	1.01-2.42
SK side-stripper steam (tons/h)	2.24-3.24
LGO side-stripper steam (tons/h)	0.66-0.90
HGO side-stripper steam (tons/h)	0.78
Top temperature (°C)	115.16-129.46
Flash zone pr. (psia)	56.94-62.63

* The temperatures are read off from the original laboratory feed TBP curves.

Network Architecture:

A network with 1 hidden layer, having a single slab, was found to be suitable for this problem. Fig. 3.4 shows the architecture used- the input layer has not been shown in the figure.

Input layer: Number of neurons: 18

Hidden layer: Number of neurons: 10

Activation function: 'tansig'

Output layer: Number of neurons: 5

Activation function: 'purelin'

The training function and the performance function used are 'trainlm' (Levenberg-Marquardt method) and 'mse' (mean square error) respectively, as in the previous model.

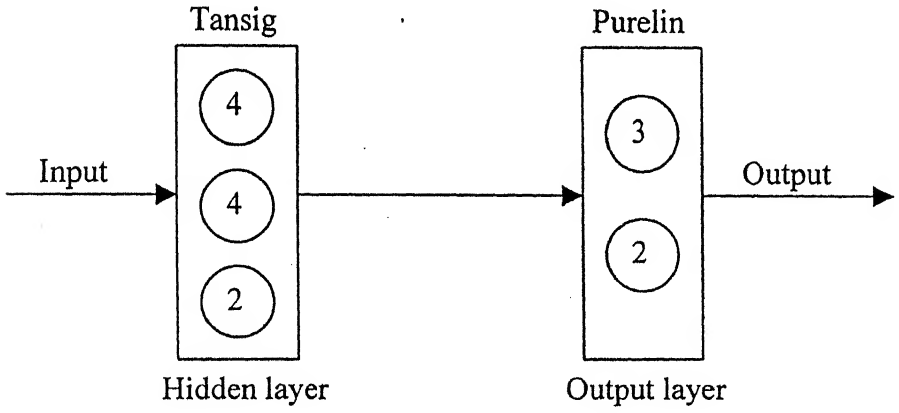


Figure 3.4 Network for predicting crude TBP curve (new model)

Normalization and Principal Component Analysis

Normalization: Neural network training can be more efficient if certain preprocessing steps are performed on the network inputs and targets. The Function 'prestd' normalizes the inputs and targets so that they will have zero mean and unity standard deviation. After the network has been trained, these vectors should be used to transform any future input, which are applied to the network. The function 'poststd' transforms the normalized data back into the original form.

Principal Component Analysis: In some situations the dimension of the input vector is large, but the components of the vector are highly correlated (redundant). In such cases it is useful to reduce the dimensions of input vector. An effective procedure for performing this operation is Principal Component Analysis. This technique has three

effects; it orthogonalizes the components of the input vector (so that they are un-correlated with each other); it orders the resulting orthogonal components so that those with the largest variation come first; and it eliminates those components which contribute the least to the variation in the data set. Depending upon the choice of variance (0.001 used in the present case), the function 'Prepca' will eliminate those principle components, which contribute less than 0.1% to the total variation in the data set. For the present study the original number of input components was reduced from 22 to 18 by using the principle component analysis.

Learning Rate & Momentum Factor:

These two parameters were chosen very carefully. The systematic study of the nature of the output versus input curves revealed that data available to us were very noisy. A low learning rate of 0.05 and a high momentum factor of 0.5 were experimented upon and a quick convergence was achieved with good generalization. The Matlab program for the model is listed in Appendix-4.

3.3 Results and Discussions

Three different types of crudes were used in the present study, one from Persian Gulf source (designated as crude-1), one from Indian Offshore field (designated as crude-2) and the third from Nigerian source (designated as crude-3). Online data were collected from an operating refinery for this investigation i.e. the different operating conditions of the CDU were recorded on-line for several days and these data were then visually screened for gross errors. Eighty-seven different sets of data were selected for further use. Fifty-one of these sets were used for training the neural net, nineteen were used for cross-validation purpose and the remaining seventeen were kept for testing the neural net developed. Besides the different operating conditions grouped together, the five TBP temperatures, read off from the original laboratory TBP curve, were included in the input data set. The first TBP temperature corresponds to the overhead distillate leaving the top of the column. The product withdrawal rates from the four side-strippers provide four more points. The outputs (target TBP temperatures) also correspond to the same volume percent points, but the temperatures are calculated by Murtuza's backcalculation procedure (Murtuza, 1999). Hence, the values of these volume percents distilled are not included in the data set.

The performance of the trained net on the test set is discussed here. Table 3.2 gives the initial TBP data for all the three crudes. Table 3.3 lists the test set having

seventeen different patterns of operating conditions, though not in their grouped forms and also the five input temperatures, along with the corresponding volume percentages distilled (the volume percentages are not fed as inputs). The ANN predicted outputs, along with the corresponding values "backcalculated" using Murtuza's procedure and the volume percentages distilled, are listed in Table 3.4. Table 3.5 lists the average and absolute deviations, absolute and percentage, as well as the two statistical parameters – Rsquared, coefficient of multiple determination, and r, the correlation coefficient, for the predictions of all the five temperatures. The terms R-squared and r are explained in Appendix-2. Table 3.6 shows the same comparison after extrapolation of the computed data to cover the entire range. Figures 3.5.1, 3.5.2 and 3.5.3 bring out the comparison in plots of the net-predicted TBP curve and the "back-calculated" TBP curve for three sets of data in the training set: data set with Serial No.1 (crude of type 1), data set with Serial No.5 (crude of type 2) and data set with Serial No. 15 (crude of type 3).

Table 3.2: TBP data for the three crudes from archives

Vol. (%)	TBP (K) Crude 1	TBP (K) Crude 2	TBP (K) Crude 3
0.0	269.66	271.56	269.66
3.00	298.72	293.71	298.72
6.36	327.35	314.33	323.47
10.82	353.46	337.24	350.52
16.36	385.68	362.24	380.24
23.00	425.66	393.21	413.35
30.73	467.90	435.72	446.93
39.55	515.95	480.35	490.07
49.46	565.66	525.04	530.04
60.46	618.17	574.78	577.78
72.55	679.70	629.70	632.70
85.73	759.70	709.70	709.70
92.73	809.59	769.59	769.59
99.99	890.78	844.79	844.79
Specific gravity	0.8712	0.8240	0.8420

Table 3.3: Operating conditions and the five input initial TBP temperatures

Table 3.3(a) Input data: Initial TBP temperatures (K)

Sl No.	Vol %	Temp 1	Vol %	Temp 2	Vol %	Temp 3	Vol %	Temp 4	Vol %	Temp 5
1	19.13	402.25	24.36	433.42	34.49	491.75	49.73	571.35	53.90	591.98
2	18.81	403.90	23.96	431.08	33.46	486.05	49.14	568.44	56.43	604.49
3	19.77	405.93	24.65	434.98	35.43	496.83	52.05	582.75	56.25	603.44
4	24.85	403.81	30.07	432.27	48.02	523.07	64.24	596.40	68.14	615.27
5	24.95	404.31	31.58	440.45	45.98	513.59	62.73	589.14	69.20	610.12
6	26.04	409.97	31.84	441.66	49.19	528.15	65.73	603.25	71.57	632.42
7	23.93	399.07	30.03	432.16	46.59	516.60	62.82	589.78	68.44	616.78
8	22.75	411.09	29.06	441.97	47.46	525.70	65.82	591.85	70.00	611.25
9	20.14	408.06	25.11	437.65	36.20	500.94	50.84	576.71	56.90	606.60
10	20.22	408.51	25.91	442.35	35.62	497.81	50.59	575.46	57.91	611.56
11	20.10	407.82	24.79	435.75	34.84	493.50	49.38	569.46	56.02	602.29
12	20.39	409.53	25.48	439.80	35.87	499.15	51.09	577.91	56.77	605.95
13	24.23	405.70	29.82	430.97	48.46	525.09	62.91	590.13	67.53	612.21
14	24.01	399.49	29.95	431.68	48.69	526.15	63.34	592.15	66.60	607.75
15	22.04	407.73	27.70	435.52	45.15	515.75	62.89	593.83	67.23	614.08
16	22.09	407.96	28.19	437.88	42.80	505.39	60.43	582.61	66.86	612.23
17	22.14	408.19	28.20	437.91	42.08	502.23	58.99	576.17	64.15	599.59

Table 3.3(b) Input data: Operating conditions

Serial No.	Crd. type	Feed flow (m3/hr)	COT (°C)	Reflux flow (m3/h)	Reflux temp. (°C)	Bottom steam (ton/h)	* Cond -ensor temp. (°C)	Flash zone pr. (Kg/sq - cm)
1	1	1251.6	346.18	200.00	31.01	6.00	73.76	3.10
2	1	1260.1	354.10	267.00	33.19	6.00	73.64	3.07
3	1	1251.10	349.74	214.00	26.10	6.01	69.66	3.08
4	2	1244.80	328.16	268.00	32.81	8.12	86.69	3.12
5	2	1282.90	329.33	233.00	35.52	7.02	82.92	3.12
6	2	1233.70	331.09	281.00	35.99	7.50	84.05	3.09
7	2	1262.80	328.30	263.00	35.12	7.50	86.13	3.07
8	3	1199.00	333.49	253.00	38.25	9.96	87.5	2.97
9	1	1237.80	354.93	240.00	39.77	8.04	90.0	3.06
10	1	1232.50	352.43	241.00	35.07	9.96	90.0	2.97
11	1	1240.70	353.31	225.00	38.45	8.56	90.0	3.10
12	1	1220.70	355.10	232.00	39.64	7.95	90.0	3.09
13	2	1218.10	330.61	264.00	43.64	8.51	90.0	3.05
14	2	1229.80	334.06	262.00	38.33	9.95	90.0	3.01
15	3	1240.70	335.35	273.00	42.87	8.04	87.5	3.07
16	3	1237.70	332.95	246.00	40.51	8.84	87.5	3.08
17	3	1234.80	330.07	234.00	39.47	8.33	87.5	3.11

*condenser outlet temperatures were not available for the'98 data set and were taken to be fixed at 90 deg C for distillation of crudes of type1 and type2 , and 87.5 deg C for crude of type 3.

Table 3.3(b) Input data:Operating conditions continued

Sl. No.	HN-PA Flow	HN-PA Temperature		LGO-PA Flow	LGO – PA Temp.		HGO-PA Flow	HGO – PA Temperature	
	(m3/h)	draw (°C)	return (°C)	(m3/h)	draw (°C)	return (°C)	(m3/h)	draw (°C)	return (°C)
1	461.60	163.23	117.10	384.30	219.11	104.12	212.10	274.85	118.94
2	599.82	165.08	103.94	374.80	210.72	99.30	263.64	267.18	124.96
3	450.06	165.62	107.25	387.60	215.65	101.00	300.28	269.63	130.21
4	609.73	167.23	99.29	374.80	222.71	95.35	300.91	264.61	127.86
5	509.78	171.89	113.37	406.00	219.75	104.21	322.03	268.45	132.23
6	479.70	168.83	103.95	380.10	219.39	96.68	319.43	265.53	131.27
7	550.14	167.21	105.67	370.30	220.28	95.98	319.05	264.27	129.42
8	746.21	161.31	126.80	445.50	225.13	129.33	430.01	283.06	165.87
9	565.69	164.01	127.05	454.60	217.62	131.01	423.30	282.82	172.79
10	410.14	161.51	115.95	445.40	218.10	126.44	412.62	277.47	164.65
11	489.42	159.75	119.05	446.90	214.73	125.66	423.62	279.39	164.66
12	489.44	162.15	121.68	445.20	216.83	129.10	415.69	282.86	169.86
13	795.39	162.25	125.97	445.10	221.28	123.36	424.08	279.08	171.11
14	719.72	166.94	127.20	420.30	225.46	118.17	425.92	283.44	163.62
15	675.67	164.40	124.08	448.70	225.12	126.85	426.21	284.78	168.92
16	541.58	170.30	125.74	446.30	226.45	130.13	432.42	281.72	168.66
17	715.14	166.12	129.88	439.80	218.76	126.87	415.57	277.78	163.25

Table 3.3(b) Input data: Operating conditions continued

Serial No.	HN-flow	Kero-flow	LGO-flow	HGO-flow	HN-SS steam	Kero-SS steam	LGO-SS Steam	HGO-SS steam
	(m3/h)	(m3/h)	(m3/h)	(m3/h)	(ton/h)	(ton/h)	(ton/h)	(ton/h)
1	65.45	126.74	190.77	52.19	2.42	2.9782	0.6743	0.78
2	64.88	119.69	197.60	91.91	1.33	2.5000	0.8946	0.78
3	61.02	134.95	207.87	52.51	1.81	3.1320	0.8764	0.78
4	65.00	223.41	201.93	48.50	1.90	3.1016	0.8729	0.78
5	85.09	184.69	214.85	83.02	1.99	2.4475	0.8843	0.78
6	71.56	214.10	204.03	72.02	2.02	3.0086	0.8838	0.78
7	76.98	209.14	204.96	70.92	2.02	3.0090	0.8915	0.78
8	75.68	220.70	220.03	50.18	1.59	2.8939	0.6735	0.78
9	61.52	137.24	181.28	75.01	1.27	3.2042	0.6721	0.78
10	70.15	119.76	184.46	90.24	1.60	2.2421	0.6747	0.78
11	58.20	124.70	180.45	82.39	1.63	2.6982	0.6697	0.78
12	62.10	126.85	185.72	69.38	1.59	2.7347	0.6700	0.78
13	68.19	227.03	176.01	56.20	1.33	3.2389	0.6690	0.78
14	72.93	230.53	180.11	40.19	1.60	2.4799	0.6736	0.78
15	70.19	216.53	220.05	53.88	1.29	3.0052	0.6732	0.78
16	75.49	180.75	218.21	79.61	1.01	2.3643	0.6709	0.78
17	74.80	171.40	208.79	63.69	1.62	2.6947	0.6718	0.78

Table 3.3(b) Input data: Operating conditions continued

Serial No.	Top temperature	HN-draw temperature	Kero-draw temperature	LGO-draw temperature	HGO-draw temperature
	(°C)	(°C)	(°C)	(°C)	(°C)
1	120.95	151.67	203.91	274.85	326.22
2	118.00	154.49	198.84	267.18	333.21
3	118.20	156.81	197.96	269.63	333.38
4	128.99	158.86	204.48	264.61	319.81
5	127.53	162.66	206.33	268.45	319.72
6	127.47	159.60	203.82	265.53	322.47
7	127.80	158.71	204.21	264.27	320.31
8	123.37	161.31	200.12	283.06	323.26
9	117.94	152.12	196.04	282.82	338.91
10	115.38	150.54	195.18	277.47	333.04
11	118.21	149.61	194.20	279.39	333.94
12	117.75	151.56	196.70	282.86	338.24
13	127.21	154.16	197.97	279.08	318.56
14	126.84	157.95	204.30	283.44	323.36
15	123.92	154.45	200.93	284.78	326.83
16	129.46	161.14	205.51	281.72	322.02
17	126.98	156.05	201.33	277.78	318.59

Table 3.4: Comparison of ANN predicted output temperatures with those obtained by backcalculation using the method of Murtuza

Sl. No	Temperature-1 (K)			Temperature-2 (K)		
	Vol.% distilled	Backcalculated Temperature	ANN predicted Temperature	Vol. % distilled	Backcalculated Temperature	ANN predicted Temperature
1	19.13	403.03	403.52	24.36	433.53	428.48
2	18.81	393.52	388.72	23.96	405.63	414.60
3	19.77	395.97	399.23	24.65	419.64	421.61
4	24.85	392.96	395.52	30.07	422.34	423.35
5	24.95	396.18	399.58	31.58	442.63	442.55
6	26.04	392.42	395.47	31.84	428.15	423.82
7	23.93	394.01	392.93	30.03	430.28	430.41
8	22.75	398.35	398.10	29.06	432.96	420.87
9	20.14	399.14	398.78	25.11	421.00	417.64
10	20.22	397.85	397.13	25.91	416.78	427.18
11	20.10	399.15	401.67	24.79	423.49	422.63
12	20.39	398.61	398.94	25.48	426.93	422.48
13	24.23	394.55	394.30	29.82	424.57	422.94
14	24.01	395.84	393.77	29.95	430.93	432.90
15	22.04	399.88	403.91	27.70	422.35	421.28
16	22.09	406.45	406.08	28.19	430.84	437.14
17	22.14	402.38	404.20	28.20	434.62	434.01

Table 3.4 continued

Sl. No	Temperature-3 (K)			Temperature-4 (K)		
	Vol.% distilled	Backcalculated Temperature	ANN predicted Temperature	Vol. % distilled	Backcalculated Temperature	ANN predicted Temperature
1	34.49	495.07	486.01	49.73	569.36	575.95
2	33.46	484.62	488.46	49.14	550.88	551.52
3	35.43	493.33	495.43	52.05	565.67	564.52
4	48.02	510.15	509.80	64.24	568.02	569.96
5	45.98	511.91	507.38	62.73	578.14	574.16
6	49.19	509.86	518.32	65.73	568.96	576.16
7	46.59	509.10	515.72	62.82	568.13	574.88
8	47.46	520.50	516.68	65.82	577.40	579.53
9	36.20	497.38	495.37	50.84	565.14	560.62
10	35.62	502.05	503.37	50.59	560.78	569.86
11	34.84	486.20	491.59	49.38	559.28	559.93
12	35.87	492.83	492.28	51.09	566.82	566.28
13	48.46	514.62	517.03	62.91	580.48	578.51
14	48.69	524.83	520.17	63.34	588.44	582.66
15	45.15	507.71	502.05	62.89	578.97	576.79
16	42.80	504.67	518.22	60.43	576.58	585.30
17	42.08	502.47	500.00	58.99	575.34	576.07

Table 3.4 continued

Serial No.	Temperature-5 (K)		
	Vol % distilled	Backcalculated Temperature	ANN predicted Temperature
1	53.90	588.65	592.77
2	56.43	602.69	586.54
3	56.25	596.81	599.34
4	68.14	597.67	602.08
5	69.20	602.56	607.90
6	71.57	605.16	612.66
7	68.44	603.97	606.30
8	70.00	604.21	606.94
9	56.90	600.14	596.49
10	57.91	605.73	607.42
11	56.02	592.71	595.32
12	56.77	594.25	595.41
13	67.53	601.14	604.06
14	66.60	603.37	600.36
15	67.23	598.26	595.66
16	66.86	604.12	614.69
17	64.15	601.78	600.34

Table 3.5: A summary of the deviations and statistical parameters for prediction of the feed TBP temperatures corresponding to percent vaporizations at the five different locations of the distillation column. (absolute deviations are in K)

Feed TBP temperatures at	Average Deviation		Maximum Deviation		R-squared value	Correlation coefficient r
	Absolute	Absolute Percentage	Absolute	Absolute Percentage		
Top tray	1.84	0.47	4.80	1.22	0.866	0.750
HN draw tray	3.78	0.89	12.09	2.79	0.773	0.598
SK/ATF draw tray	4.52	0.90	13.55	2.69	0.876	0.767
LGO draw tray	3.80	0.67	9.08	1.62	0.865	0.748
HGO draw tray	4.40	0.73	16.15	2.68	0.622	0.387

The five temperatures predicted by the net, along with the IBP, TBP 80% and FBP points taken from the original TBP curve, can be joined by straight lines to give what is the extrapolated complete TBP curve. Table 3.4 shows the comparison between the two extrapolated TBP curves -one obtained by joining the five points predicted by the ANN, the other one obtained by joining the five points "backcalculated" by Murtuza's procedure. Since the points from TBP 80% onwards will be the same for both the cases, those points are not shown in the table.

Table 3.6: Comparison of ANN predicted TBP with backcalculated TBP

Sl. No.	Temperature-1 (Volume-5%)			Temperature-2 (Volume-10%)		
	Backcalculated Temp. (K)	ANN predicted Temp. (K)	error (K)	Backcalculated Temp. (K)	ANN predicted Temp. (K)	Error (K)
1	304.52	304.78	-0.26	339.37	339.90	-0.53
2	302.58	302.01	0.57	335.51	334.35	1.16
3	301.61	302.89	–1.28	333.55	336.12	-2.57
4	295.98	296.71	-0.73	320.41	321.87	-1.46
5	296.53	297.22	-0.69	321.51	322.89	-1.38
6	294.77	295.50	-0.73	317.98	319.45	-1.47
7	297.14	297.58	-0.44	322.72	323.61	-0.89
8	297.95	300.48	-2.53	326.24	331.30	-5.06
9	301.81	302.34	-0.53	333.96	335.02	-1.06
10	301.36	301.80	-0.44	333.07	333.94	-0.87
11	301.88	302.83	-0.95	334.10	336.00	-1.90
12	301.28	301.91	-0.63	332.89	334.16	-1.27
13	296.94	297.10	-0.16	322.33	322.64	-0.31
14	297.44	297.63	-0.19	323.31	323.71	-0.40
15	299.20	300.34	-1.14	328.73	331.01	-2.28
16	300.62	301.16	-0.54	331.57	332.66	-1.09
17	299.63	300.28	-0.65	329.60	330.91	-1.31

Table 3.6 continued

Sl. No.	Temperature-3 (Volume-20%)			Temperature-4 (Volume-30%)		
	Backcalculated Temp. (K)	ANN predicted Temp. (K)	error (K)	Backcalculated Temp. (K)	ANN predicted Temp. (K)	error (K)
1	407.94	408.33	-0.39	465.75	461.54	4.21
2	396.23	395.80	0.43	453.02	457.59	-4.57
3	397.06	402.26	-5.20	454.07	460.79	-6.72
4	369.25	372.18	-2.93	420.95	425.49	-4.54
5	371.46	374.22	-2.76	430.42	431.99	-1.57
6	364.40	367.34	-2.94	416.05	413.30	2.75
7	373.88	375.66	-1.78	428.93	429.98	-1.05
8	382.82	392.93	-10.11	436.36	425.96	10.40
9	398.26	400.39	-2.13	452.62	453.28	-0.66
10	396.48	398.22	-1.74	450.06	457.79	-7.73
11	398.54	402.34	-3.80	454.12	457.45	-3.33
12	396.13	398.65	-2.52	453.72	454.43	-0.71
13	373.10	373.72	-0.62	424.51	423.97	0.54
14	375.06	375.85	-0.79	430.10	434.01	-3.91
15	387.80	392.37	-4.57	432.37	430.84	1.53
16	393.49	395.66	-2.17	438.73	444.16	-5.43
17	389.55	392.16	-2.61	442.22	440.57	1.65

Table 3.6 continued

Sl. No.	Temperature-5 (Volume-40%)			Temperature-6 (Volume-50%)		
	Backcalculated Temp. (K)	ANN predicted Temp. (K)	error (K)	Backcalculated Temp. (K)	ANN predicted Temp. (K)	Error (K)
1	518.86	519.15	-0.29	566.22	577.47	-11.25
2	509.57	510.86	-1.29	550.61	551.87	-1.26
3	510.57	515.02	-4.45	552.87	557.92	-5.05
4	468.52	472.41	-3.89	514.36	517.25	-2.89
5	480.87	478.85	2.02	524.75	521.36	3.39
6	464.51	466.06	-1.55	510.07	519.37	-9.30
7	475.32	479.92	-4.60	518.54	525.63	-7.09
8	482.48	470.19	12.29	526.96	513.82	13.14
9	512.22	510.62	1.60	557.17	556.75	0.42
10	516.89	519.15	-2.26	555.01	563.72	-8.71
11	509.13	514.01	-4.88	557.94	561.47	-3.53
12	510.03	511.54	-1.51	557.28	560.99	-3.71
13	471.35	474.12	-2.77	517.98	523.03	-5.05
14	478.78	479.30	-0.52	527.01	524. 28	2.73
15	479.79	477.58	2.21	523.76	521.55	2.21
16	487.75	498.92	-11.17	530.61	541.61	-11.00
17	489.61	488.28	1.33	532.97	532.86	0.11

Table 3.6 continued

Sl. No.	Temperature-7 (Volume-60%)			Temperature-8 (Volume-75%)		
	Backcalculated Temp. (K)	ANN predicted Temp. (K)	error (K)	Backcalculated Temp. (K)	ANN predicted Temp. (K)	Error (K)
1	615.13	622.15	-7.02	693.78	695.54	-1.76
2	615.42	603.78	11.64	693.85	690.94	2.91
3	611.24	617.12	-5.88	692.81	694.28	-1.47
4	548.99	553.92	-4.93	635.53	641.19	-5.66
5	563.15	562.21	0.94	640.85	639.68	1.17
6	544.76	553.95	-9.19	623.81	636.01	-12.20
7	553.85	563.20	-9.35	637.56	641.96	-4.40
8	563.77	555.73	8.04	642.57	637.23	5.34
9	610.99	607.92	3.07	692.75	691.98	0.77
10	610.94	613.21	-2.27	692.73	693.30	-0.57
11	608.80	611.96	-3.16	692.20	692.99	-0.79
12	606.34	609.13	-2.79	691.59	692.28	-0.69
13	562.23	565.14	-2.91	639.06	643.66	-4.60
14	569.18	567.14	2.04	642.39	644.01	-1.62
15	562.78	562.57	0.21	638.48	640.71	-2.23
16	570.22	579.62	-9.40	642.04	647.96	-5.92
17	574.81	577.16	-2.35	646.58	647.15	-0.57

From the tables (Tables 3.4 and 3.6) as well as the graphs (Figures 3.5.1, 3.5.2 and 3.5.3) it can be seen that the ANN predicted values match reasonably well with the values predicted by the backcalculation procedure of Murtuza. The deviation of ANN predicted TBP temperatures from the corresponding "backcalculated" values lies within ± 5 °C most of the time. However, in a few cases the deviation exceeds 10 °C. The statistical parameters listed in Table 3.5 also do not have very satisfactory values. One reason for poor prediction in these cases could be the lack of sufficient number of datasets. The performance of the ANN model depends on how well it has been trained, which, in turn, depends on how many data sets it has been trained with. Here, the number of data sets for training is only 51 and hence the exact nature of relationships between the outputs and the inputs might not have been fully represented in so few data sets, resulting in poor prediction in certain cases. However, better prediction is expected when the net is retrained with more number of data sets. While there is no rule that specifies the minimum number of patterns needed to train a net,

Figure 3.5.1 Comparison of ANN predicted TBP curve with the TBP curve generated by the 'backcalculation' procedure of Murtuza for crude 1.

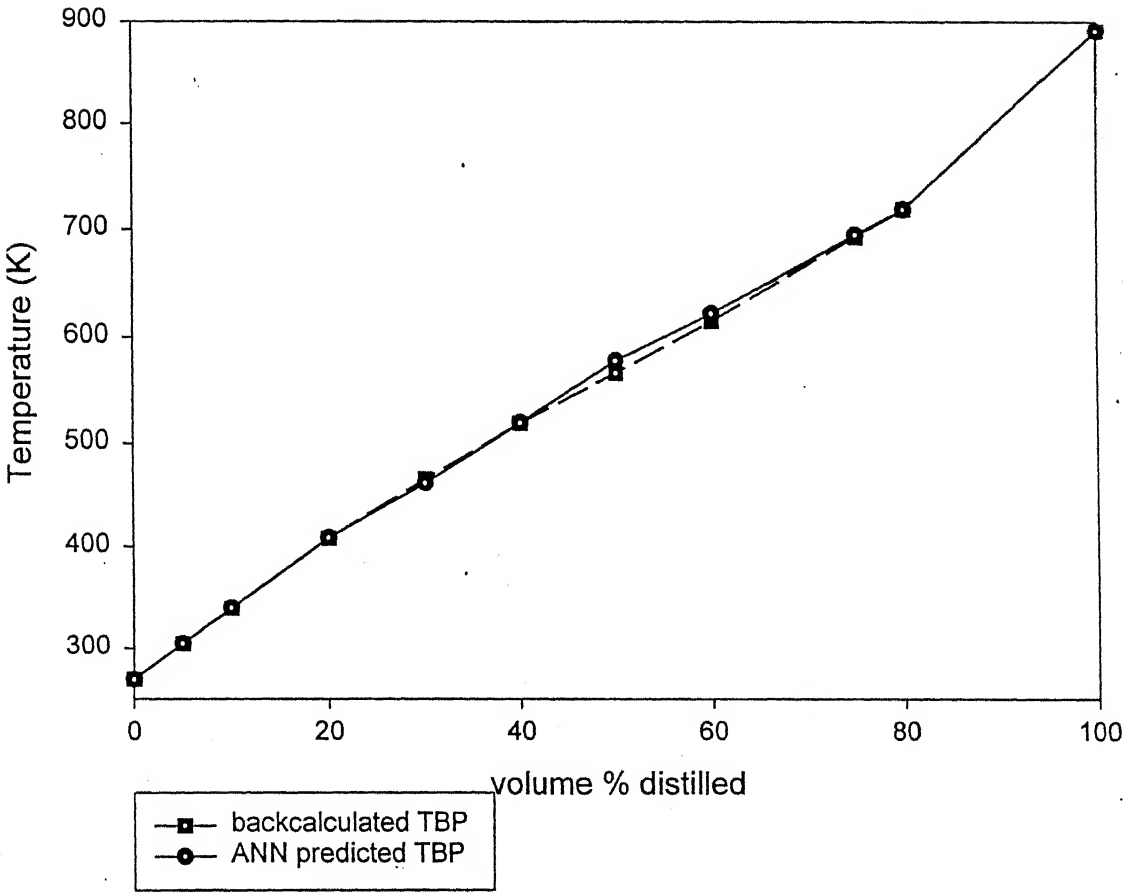


Figure 3.5.2 Comparison of ANN predicted TBP curve with the TBP curve generated by the 'backcalculation' procedure of Murtuza for crude 2.

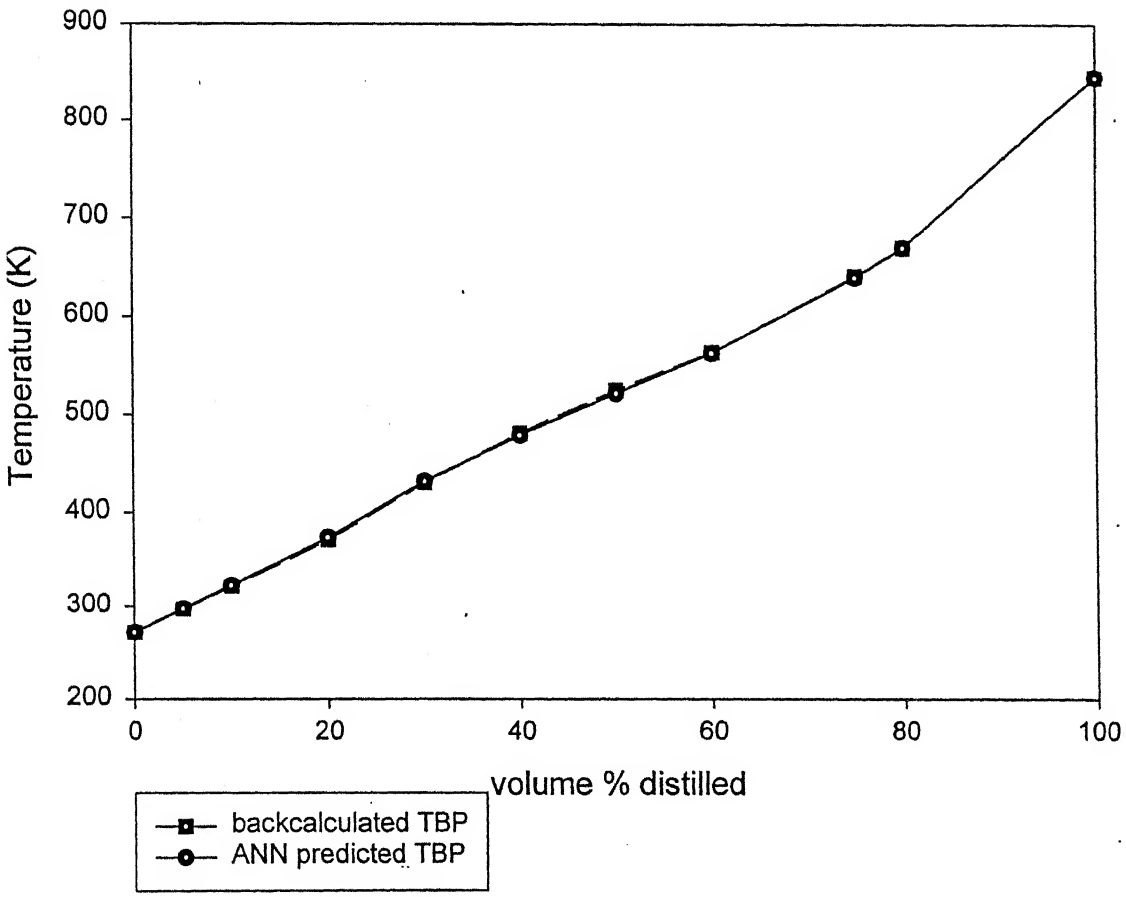
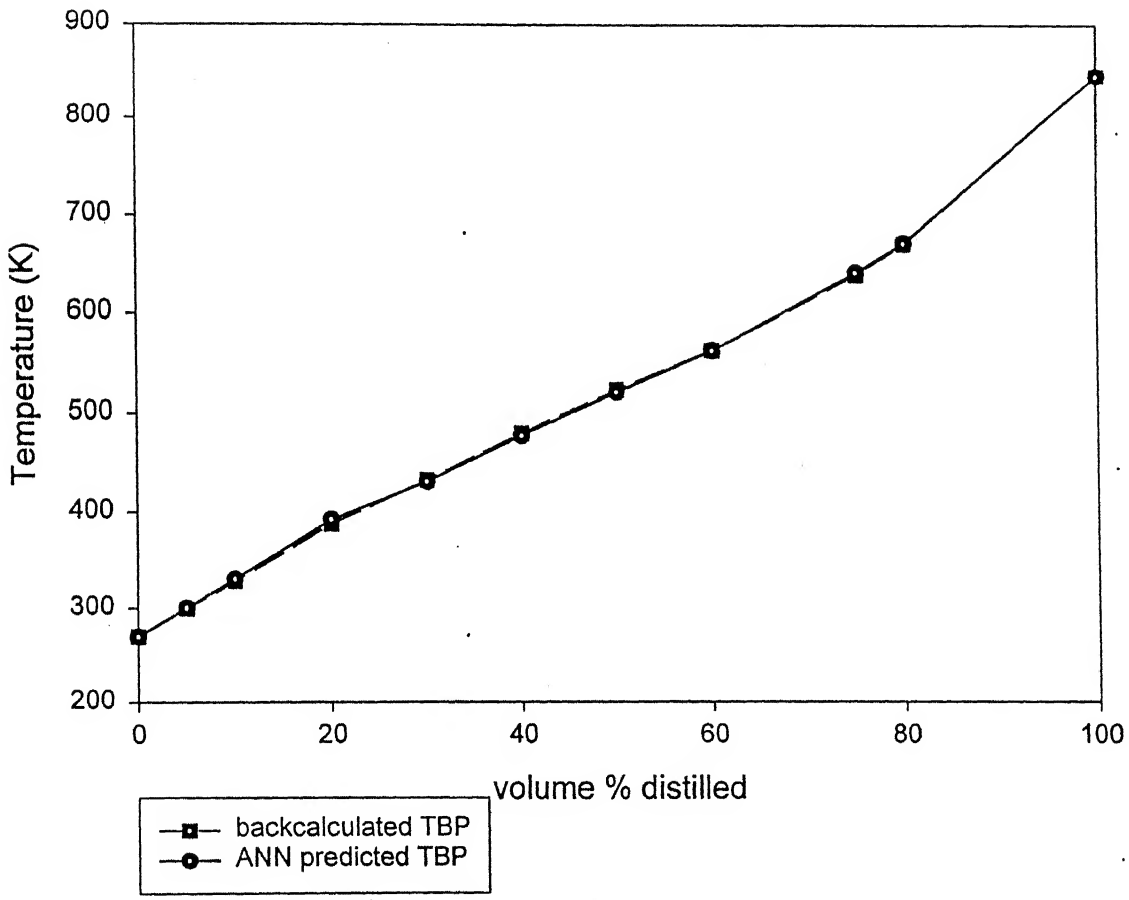


Figure 3.5.3 Comparison of ANN predicted TBP curve with the TBP curve generated by the 'backcalculation' procedure of Murtuza for crude 3.



the two thumb rules, discussed in section 2.6 of the previous chapter, can be followed to ensure that the net is trained to generalize well enough.

Chapter 4

ANN models for prediction of properties of Petroleum fractions

This chapter discusses the development of models for the prediction of ASTM temperatures of two important petroleum fractions - Heavy Naphtha (HN) and Superior Kerosene/Aviation Turbine Fuel (SK/ATF) - and the Flash Point for SK/ATF. There are no hardware sensors available for the measurement of these properties online. Usually samples are collected once every eight hours and the properties are measured in the laboratory. It is, therefore, not possible to achieve proper quality control through feedback mechanism. In the present work, an attempt has been made to predict these properties online by correlating them with variables that are measurable online i.e. the operating conditions, specified in terms of various temperatures, pressures and flow rates. Different ANN based models have been developed which predict the above mentioned product properties, when supplied with sets of input variables. As mentioned earlier, the product properties depend not only on the operating conditions but on the characteristics of the crude oil being processed as well. So the different points of the TBP curve of the crude, predicted by the ANN model described in the previous chapter, also form a part of the set of input variables to the neural nets. The different operating conditions are the other input variables. Thus the neural net for online prediction of TBP curve, connected in series with the nets for property prediction, forms the package for online estimation of product properties. This chapter, however, discusses only the nets for property prediction i.e. prediction of ASTM temperatures of HN, SK/ATF and the flash point of SK/ATF. The net for online TBP prediction, discussed in the previous chapter, will henceforth be referred to as Net-1 for the sake of convenience.

4.1 ASTM Temperatures

The ubiquity of the use of ASTM distillation to characterize the composition of petroleum fractions makes the prediction of ASTM boiling curve of a product fraction extremely important. ASTM distillation is a rapid procedure used for analysis of petroleum products and intermediate fractions. This procedure was developed by the American Society for Testing Materials and hence the name. It is basically a rapid batch distillation run in Engler flask, employing no trays or reflux between the stillpot and the condenser. The only reflux available is that generated by heat losses from the neck of the flask. This test method is used in control laboratories throughout the world. ASTM temperature curve is a plot of ASTM boiling points vs. volume percent distilled, very

similar to TBP curve. A detailed discussion of the procedures involved is available in the literature (API Technical Data Book, 1982).

While predicting ASTM temperatures of different products through ANN, the input variables will be different for different products. Even for the same product, the temperatures on the upper part of the ASTM curve depend on certain variables, but the same variables do not affect the values on the lower part as much. For example, the amount of steam for the SK side stripper is an important input variable for determining the upper part of the ASTM curve of HN (>ASTM 50%) and the lower part of the ASTM curve of SK/ATF (<ASTM 50%) but it hardly affects the values on the lower part of the ASTM curve of HN or on the upper part of the ASTM curve of SK/ATF. Thus, by simply considering the physics of the problem, we can eliminate certain variables while predicting a particular ASTM temperature. Hence, several neural nets, instead of one single net, have been built to predict the different ASTM temperatures. There are four different nets to calculate the various ASTM temperatures, clubbed together in groups of two or three in the following way-**1)** Initial Boiling Point (IBP), ASTM 5% and ASTM 10% for HN **2)** ASTM 90%, ASTM 95% and FBP for HN **3)** IBP and ASTM 5% for SK/ATF and **4)** ASTM 95% and FBP for SK/ATF. These four nets for calculating these temperatures are discussed in the following subsections.

4.1.1 ANN Model for Predicting IBP, ASTM 5% and ASTM 10% Temperatures for Heavy Naphtha (Net-2a)

Inputs to the ANN model: The input variables for the neural net were chosen carefully, taking into account the chemical engineering aspect of the problem. For example, operating conditions like SK side-stripper steam flow rate, LGO pump around flow rate etc. that can have no effect on the IBP, ASTM 5% and ASTM 10% temperatures of HN were not fed as inputs to the net. Only those operating conditions, which were considered to be sufficiently important so as to affect the above-mentioned temperatures, were supplied as input variables to the net, clubbed in groups of twos or threes as in the case for online TBP estimation. The input variables were the following: a)TBP 5%, TBP 10%, TBP20% and TBP30% points predicted by Net-1. b) crude specific gravity c)feed flow rate × COT d)reflux flow × reflux temperature e) HN flow rate × HN draw tray temperature f) bottom steam flow rate g) HN side stripper steam flow rate h) flash zone pressure i)top temperature . Thus there are 12 input variables. The range of variation of the variables is listed in Table 4.2. Application of principal component analysis decreased the number of variables to 10 and thus the input layer has 10 neurons to receive the 10 inputs.

Outputs from the ANN model: There are three neurons in the output layer to predict the three outputs from the net - IBP, ASTM 5% and the ASTM 10% temperatures of HN.

The training, validation and the prediction sets are chosen from the entire data set in the same way as in the case for TBP prediction, and hence is not discussed again in this chapter. However, for some data sets the laboratory measured properties were not available, resulting in the exclusion of these data sets from the model altogether.

The outputs for the training and the validation sets are the three laboratory measured ASTM temperatures of HN, produced from the CDU under the operating conditions and crude type which form the corresponding set of input variables.

Network Architecture: The network architecture was decided upon by a trial and error procedure. After experimenting with several architectures, the one, which was found to perform best, had 2 slabs in the hidden layer, connected in series. The final network structure, also outlined in Table 4.1, is as follows:

Input layer: Number of neurons: - 10.

Hidden layer:

Slab-1: Number of neurons: -3

Activation function: 'logsig'.

Slab-2: Number of neurons: - 8

Activation function: 'tansig'

Output layer: Number neurons: - 3

Activation function: 'purelin'

Figure 4.1 shows the architecture used, the input layer not being included in the figure.

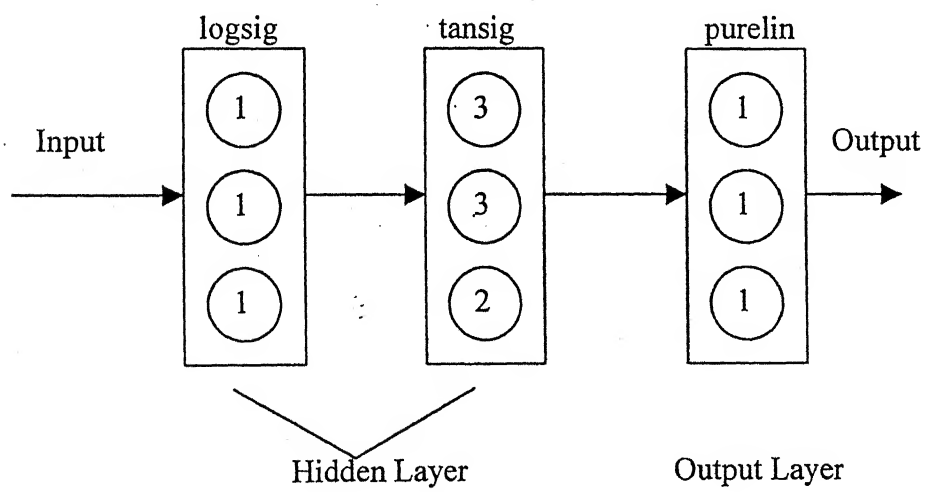


Figure 4.1 Network for predicting IBP, ASTM 5% and ASTM 10% temperatures for Heavy Naphtha

The training function and the performance functions used are 'trainlm' (Levenberg-Marquardt method) and 'mse' (mean square error) respectively.

Learning rate and momentum factor: A learning rate of 0.05 and a momentum factor of 0.5 have been used to achieve good convergence.

4.1.2 ANN Model for Predicting ASTM 90%, ASTM 95% and FBP for Heavy Naphtha (Net-2b)

The network architecture selected for this model is presented in Table 4.1 and the input variables to the net, along with the range of these variables, are given in Table 4.2. The outputs for this model are expected to be affected by variables like SK side-stripper steam flow rates, HN pumparound flow rate, the draw and return temperatures of this pumparound, the draw rate and draw temperature of the SK/ATF stream. Moreover, HN side-stripper steam flow rate will not affect the output variables. Though the total number of input variables for this model is 14 (Table 4.2), the application of Principal Component Analysis decreased the number of variables to 12 and hence the input layer contains 12 neurons. The output layer contains as many number of neurons as the number of output variables i.e three neurons. The activation function chosen for the output layer is 'purelin'. The training function and the performance functions used are 'trainlm' (Levenberg-Marquardt method) and 'mse' (mean square error) respectively.

4.1.3 ANN Model for Predicting Specific Gravity, IBP and ASTM 5% for SK/ATF (Net–2c)

Since SK/ATF is a product drawn from altogether a different tray in the distillation column, it is obvious that the ASTM temperatures of this draw are not controlled by the same operating conditions and the same temperatures on the feed TBP curve as in the case for the previous draw i.e. Heavy Naphtha. Inclusion of the TBP 40%, TBP 50%, TBP 60% and TBP 75% points from the feed TBP curve predicted by Net-1 in the model inputs is an important feature of Net-2c (Table 4.2). The operating conditions that are fed as inputs are the same as that for Net-2b. The final number of input variables after the application of Principal Component Analysis is 13 and the input layer contains the same number of neurons. The architecture for this model is also slightly complex. The hidden layer has three slabs, two of them connected in parallel with the third one in series (Table 4.1). However, the activation function for the output layer in this case too is 'purelin' and

the number of neurons in the output layer is three. The training and the performance functions also remain the same as before -'trainlm' and 'mse' respectively.

4.1.4 ANN Model for Predicting ASTM 95% and FBP for SK/ATF (Net-2d)

The outputs from this model being points on the upper part of the ASTM curve of SK/ATF are expected to be sensitive to the LGO draw rate and temperature, LGO side-stripper steam as well as the LGO pumparound flow rates, the draw and return temperatures of the pumparound. Inclusion of several new input variables, pertaining to the LGO stream, is an interesting feature of Net-2d. There are 18 input variables, as listed in Table 4.2, but application of Principal Component Analysis decreased the number to 13. The network (Table 4.1) consists of two slabs in the hidden layer, connected in series. The input and output layers are designed in the line of the other nets. The training and the performance functions remain 'trainlm' and 'mse' respectively.

Typical convergence graphs for training the neural nets are shown in Appendix-3. These are plots of the sum of square of errors for the training data set as well at the validation data set vs. the number of epochs.

The MATLAB programs for all the models are listed in Appendix-4.

Table 4.1: Summary of the network architecture used for various models

Model No.	Hidden Layer			Learning rate	Momentum factor
	Slab No.	No. of neurons	Activation function		
Net-2a	1	3	logsig	0.05	0.5
	2	8	tansig		
Net-2b	1	3	logsig	0.05	0.5
	2	10	tansig		
Net-2c*	1	5	radbas	0.05	0.5
	2	5	logsig		
	3	15	tansig		
Net-2d	1	5	logsig	0.05	0.5
	2	15	tansig		
Net-3	1	4	logsig	0.05	0.5
	2	12	tansig		

*The configuration for Net-2c, being slightly complex, is shown below for a better understanding.

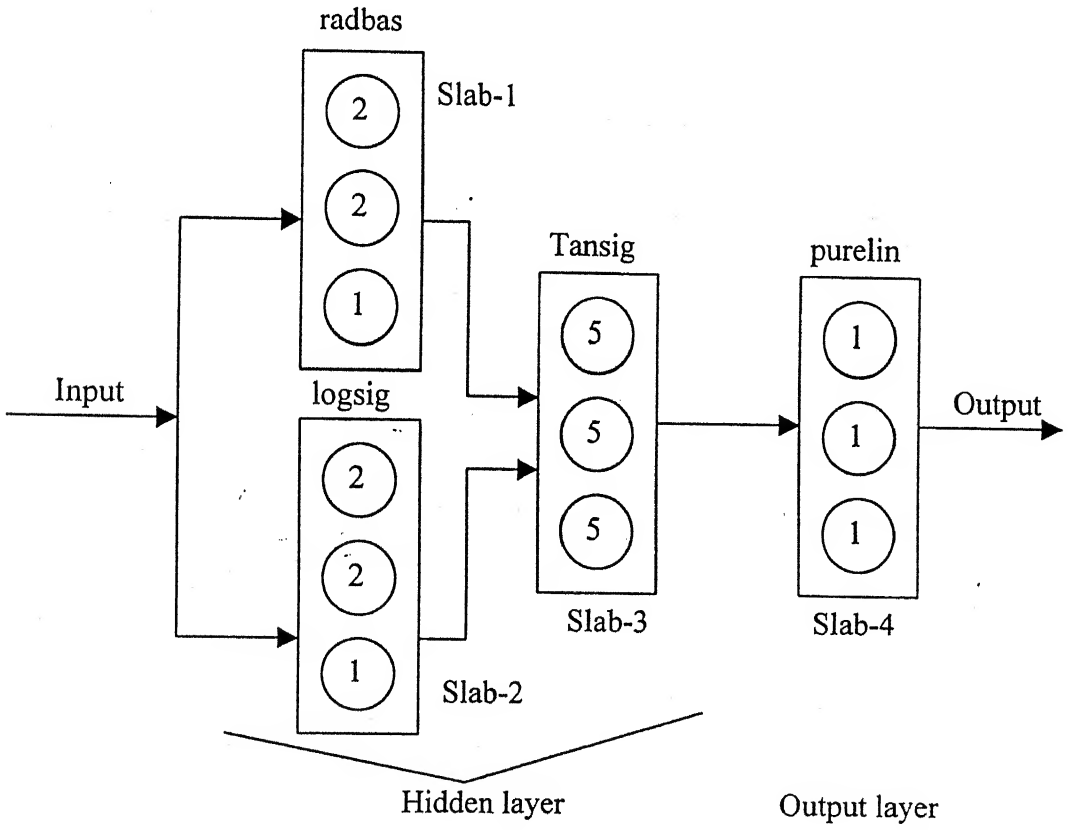


Table 4.2: Inputs for ASTM temperatures prediction and range of data used

Model Inputs	Input data range for				
	Net-2a	Net-2b	Net-2c	Net-2d	Net-3
TBP 5% temperature (K) of feed	←		294.8-304.5		→
TBP 10% temperature (K) of feed	←		317.0-339.4		→
TBP 20% temperature (K) of feed	←		363.0-407.9		→
TBP 30% temperature (K) of feed	←		412.2-466.7		→
TBP 40% temperature (K) of feed	—	—	←	460.8-524.2	→
TBP 50% temperature (K) of feed	—	—	←	508.0-566.2	→
TBP 60% temperature (K) of feed	—	—	←	544.4-617.5	→
TBP 75% temperature (K) of feed	—	—	←	623.8-694.4	→
Crude Specific Gravity	←		0.824-0.8712		→
Feed flow rate(m3/h) × COT (°C)	←		3.97×10^4-4.46×10^5		→
Reflux flow rate (m3/h) × Reflux temp. (°C)	←		5.57×10^3-1.38×10^5		→
HN flow rate (m3/h) × draw tray temp.(°C)	←		8.70×10^3- 1.40×10^4		→
SK/ATF flow rate(m3/h) × draw tray temp.(°C)	—	←	2.34×10^4-4.92×10^4		→

Table 4.2 continued

Model Inputs	Input data range for				
	Net-2a	Net-2b	Net-2c	Net-2d	Net-3
	—	—	—	4.79×10^4 - 6.27×10^4	—
HN pumparound flow rate (m3/h) × (pump-around draw temp. (°C) × return temp. (°C))	—	1.68×10^4 - 4.14×10^4		—	1.68×10^4 - 4.14×10^4
SK/ATF pumparound flow rate (m3/h) × (pump-around draw temp. (°C) × return temp. (°C))	—	—	—	3.73×10^4 - 5.15×10^4	—
LGO pumparound flow rate (m3/h) × (pump-around draw temp. (°C) × return temp. (°C))	—	—	—	3.31×10^4 - 5.10×10^4	—
Bottom steam flow rate (tons/h)	← 6.00-9.96 →				
HN side-stripper steam flow rate (tons/h)	1.01 – 2.42	—	—	—	—
SK side-stripper steam flow rate (tons/h)	—	2.24 – 3.24		—	—
LGO side-stripper steam flow rate (tons/h)	—	—	—	0.66 – 0.90	—
Top tray temperature (°C)	115.2 – 129.5		—	—	—
Flash zone pr. (psia)	← 56.94-62.63 →				

4.1.5 Results and discussions

As already stated, the usefulness of a net lies in its ability to correctly predict the output when presented with input data that it has never seen before. In this section, the performance of the four nets with respect to the predictions made on the test sets only is discussed. The operating conditions and the TBP points predicted by Net-1, which together form the set of input variables for these four nets, are not presented in this section. Only the net predicted values, and their comparison with the actual laboratory measured values are presented in a tabular form, along with the type of feed crude oil. Table 4.3a shows a comparison of ANN predicted ASTM temperatures for HN with the laboratory measured values. Table 4.3b shows the actual as well as percent differences between the net predicted values and the laboratory measured values given in Table 4.3a. Table 4.4a and Table 4.4b show similar results for the SK/ATF fraction except that the specific gravity has also been included in this case. The maximum absolute deviation and the average absolute deviation, as well as the two statistical parameters – R-squared value and r – for the different properties are summarized in Table 4.5. Figure 4.2 shows the parity plot for the different ASTM temperatures. From the tabulated results, as well as the parity plot, it is seen that the predictions from the ANN models are generally satisfactory and except for a few stray cases, the predictions are not entirely off the mark. The average absolute deviations in the case of ASTM temperatures are usually 3 to 4°C, which may be considered acceptable. But the maximum deviations in most cases are in the neighbourhood of 10°C with the largest being 20°C in case of IBP temperature for SK/ATF. However, there is an uncertainty in the measurement of IBP itself, which can account for this large deviation. It can be observed from Tables 4.3b and 4.4b that a few stray cases have large errors while most of the cases have errors close to the average. Except for the IBP of SK/ATF fraction, the average percentage deviation and the maximum percentage deviation are, for all the cases, less than ±3% and ±7.5% respectively. The prediction can be claimed to be very good for the density of SK/ATF fractions where the maximum percentage deviation is 1.75%. The statistical parameters, however, for most of the cases, have far from satisfactory values. While stray large errors is definitely one of the reasons for the unacceptable values of these parameters, the fact that the overall predictions are not very good can also not be overlooked altogether. The parameters are reasonably satisfactory for the ASTM 95% and FBP of SK/ATF. The possible reasons for the modest predictions of the nets are the insufficient data in the training set and the noise inherent in the entire data set. A better generalization can be expected once the nets are trained with more data in the training set. The data in the training set should cover the entire range within which the operating conditions might vary, so as to enable learn the exact nature of relationships between the inputs and the outputs.

Table 4.3a: ASTM temperatures of Heavy Naphtha (°C)

Sl. No.	Crd. Type	IBP		ASTM 5%		ASTM 10%		ASTM 90%		ASTM 95%		FBP	
		ANN output	Lab data	ANN output	Lab data	ANN output	Lab data	ANN output	Lab data	ANN output	Lab data	ANN output	Lab data
1	1	116.54	115	125.22	125	128.14	128	148.72	155	153.51	159	180.68	181
2	1	109.19	114	118.63	120	122.10	123	143.70	146	148.46	152	173.59	180
3	1	114.48	114	122.64	122	125.59	125	143.63	145	148.29	149	166.53	168
4	2	109.39	108	118.90	117	122.13	120	150.24	145	153.49	147	170.41	165
5	2	116.09	108	124.76	117	127.00	120	146.41	145	151.23	147	172.86	165
6	3	109.30	102	118.67	115	122.11	118	145.45	142	150.18	145	165.97	163
7	1	108.96	107	118.50	120	121.82	124	143.66	153	148.41	157	173.22	182
8	1	108.52	110	118.49	121	121.83	125	148.83	152	153.18	156	169.16	176
9	1	108.10	115	118.27	122	121.79	125	143.87	150	148.60	155	170.57	176
10	1	109.42	111	118.77	118	122.26	120	144.07	143	148.76	147	172.65	174
11	2	117.51	110	122.39	121	122.10	125	151.48	152	154.49	156	171.04	176
12	3	110.05	107	119.13	120	122.26	124	143.59	153	148.33	157	173.53	182
13	3	101.96	100	113.21	118	117.80	123	151.75	155	158.63	159	182.67	181
14	3	111.54	115	121.12	122	124.85	125	151.08	150	157.77	155	181.58	176

Table 4.3b: Differences between the ANN predictions and the laboratory measured data for ASTM temperatures of Heavy Naphtha (actual differences are in °C)

Sl. No.	Crd. Type	IBP		ASTM 5%		ASTM 10%		ASTM 90%		ASTM 95%		FBP	
		Actual	Percent	Actual	Percent	Actual	Percent	Actual	Percent	Actual	Percent	Actual	Percent
1	1	1.54	1.34	0.22	0.18	0.15	0.12	-6.28	-4.05	-5.49	-3.45	-0.32	-0.18
2	1	-4.81	- 4.22	-1.37	-1.14	-0.90	-0.73	-2.29	-1.57	-3.54	-2.33	-6.41	-3.57
3	1	0.48	0.42	0.64	0.48	0.59	0.47	-1.37	-0.95	-0.71	-0.48	-1.47	-0.88
4	2	1.39	1.29	1.90	1.62	2.13	1.78	5.25	3.62	6.49	4.41	5.41	3.28
5	2	8.09	7.49	7.76	6.63	7.00	5.83	1.41	0.97	4.23	2.88	7.86	4.76
6	3	7.30	7.16	3.67	3.19	4.11	3.48	3.45	2.43	5.18	3.57	2.97	1.82
7	1	1.96	1.83	-1.50	-1.25	-2.18	-1.76	-9.34	-6.10	-8.59	-5.47	-8.78	-4.82
8	1	-1.48	-1.35	-2.51	-2.07	-3.17	-2.54	-3.17	-2.09	-2.82	-1.81	-6.84	-3.89
9	1	-6.90	-6.00	-3.73	-3.06	-3.21	-2.57	-6.13	-4.09	-6.40	-4.13	-5.43	-3.09
10	1	-1.58	-1.42	0.77	0.65	2.26	1.88	1.07	0.75	1.76	1.20	-1.35	-0.78
11	2	7.51	6.83	1.39	1.15	-2.90	-2.32	-0.52	-0.34	-1.51	-0.97	-4.96	-2.82
12	3	3.05	2.85	-0.87	-0.73	-1.74	-1.40	-9.41	-6.15	-8.67	-5.52	-8.47	-4.65
13	3	1.96	1.96	-4.79	-4.06	-5.19	-4.22	-3.25	-2.10	-0.37	-0.23	1.67	0.92
14	3	-3.46	- 3.01	-0.88	-0.72	-0.15	-0.12	1.08	0.72	2.77	1.79	5.58	3.17

Table 4.4a: Specific Gravity and ASTM temperatures of SK/AT (all ASTM temperatures are in °C)

Sl. No.	Crd. Type	Specific Gravity		IBP		ASTM 5%		ASTM 95%		FBP	
		ANN Output	Lab Data	ANN output	Lab data	ANN output	Lab data	ANN output	Lab data	ANN output	Lab data
1	1	0.7894	0.7940	145.14	154	164.61	169	234.09	234	250.63	250
2	1	0.7876	0.7870	143.14	155	164.18	164	219.47	217	244.04	240
3	1	0.7927	0.7920	142.74	148	163.59	163	231.98	227	248.15	249
4	2	0.8040	0.8060	138.22	140	156.92	162	248.07	252	264.25	271
5	2	0.8104	0.8100	147.91	145	162.73	159	242.93	238	260.38	251
6	2	0.8024	0.8040	138.54	137	158.51	156	244.52	245	259.52	261
7	2	0.8110	0.8040	145.66	135	160.68	157	242.84	243	260.91	265
8	3	0.7976	0.8118	135.79	142	160.15	165	245.09	248	264.15	269
9	1	0.7852	0.7901	140.62	151	160.04	165	230.42	237	249.68	261
10	1	0.7919	0.7901	136.38	138	157.83	161	229.43	230	247.73	241
11	1	0.7897	0.7928	132.12	153	159.12	162	236.20	240	249.99	250
12	1	0.7849	0.7953	130.40	150	155.90	162	237.37	243	250.94	258
13	2	0.7985	0.8062	136.18	133	155.01	157	250.78	243	264.74	261
14	2	0.8138	0.8107	140.89	135	166.50	160	253.54	264	268.35	272
15	3	0.8047	0.8121	142.45	137	161.50	159	244.82	250	260.91	260
16	3	0.8003	0.8114	147.12	150	163.23	167	243.66	246	256.96	257
17	3	0.8062	0.8098	147.06	134	164.88	162	246.28	247	256.04	258

Table 4.4b: Differences between ANN predictions and laboratory measurements for properties of SK/ATF

(actual differences for the ASTM temperatures are in °C)

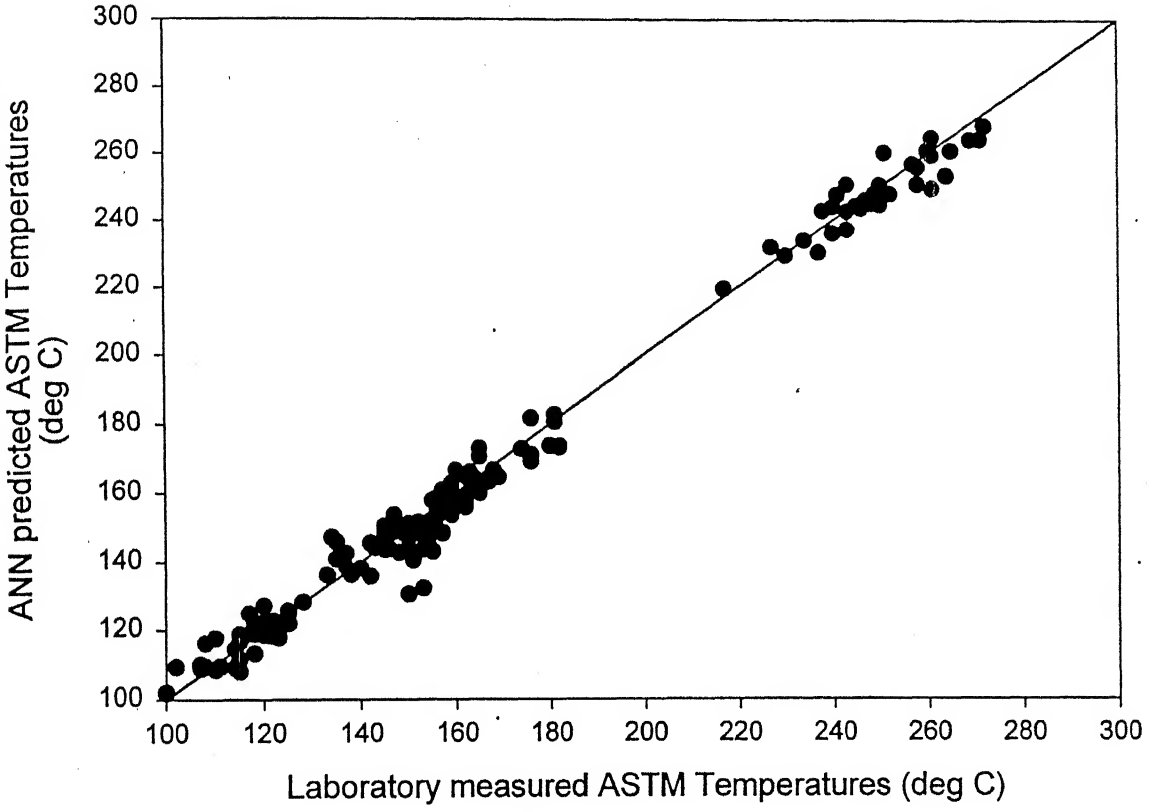
Serial No.	Crude Type	Specific gravity		IBP		ASTM 5%		ASTM 95%		FBP	
		Actual	Percent	Actual	Percent	Actual	Percent	Actual	Percent	Actual	Percent
1	1	-0.0046	-0.58	-8.86	-5.75	-4.39	-2.60	0.09	0.04	0.63	0.25
2	1	0.0006	0.08	-11.86	-7.65	0.18	0.11	2.47	1.14	4.04	1.68
3	1	0.0007	0.09	-5.26	-3.55	0.59	0.36	4.98	2.19	-0.85	-0.34
4	2	-0.0020	-0.25	-1.78	-1.27	-5.08	-3.14	-3.93	-1.56	-6.75	-2.49
5	2	0.0004	0.05	2.90	2.00	3.73	2.35	4.93	2.07	9.38	3.74
6	2	-0.0016	-0.20	1.54	1.12	2.51	1.61	-0.48	-0.20	-1.48	-0.57
7	2	0.0070	0.87	10.66	7.90	3.68	2.35	0.16	0.07	-4.09	-1.54
8	3	-0.0142	-1.75	-6.21	-4.37	-4.85	-2.94	-2.91	-1.17	-4.85	-1.80
9	1	-0.0049	-0.62	-10.38	-6.87	-4.96	-3.01	-6.58	-2.78	-11.32	-4.34
10	1	0.0018	0.23	-1.62	-1.17	-3.17	-1.97	-0.57	-0.25	6.73	2.79
11	1	-0.0031	-0.39	-20.88	-13.65	-2.88	-1.78	-3.80	-1.58	-0.01	0.00
12	1	-0.0104	-1.31	-19.60	-13.07	-6.10	-3.76	-5.63	-2.32	-7.06	-2.74
13	2	-0.0077	-0.96	3.18	2.39	-1.99	-1.27	7.78	3.20	3.74	1.43
14	2	0.0031	0.38	5.89	4.36	6.50	4.06	-10.46	-3.96	-3.65	-0.09
15	3	-0.0074	-0.91	5.45	3.98	2.50	1.57	-5.18	-2.07	0.91	0.35
16	3	-0.0111	-1.37	-2.88	-1.92	-3.77	-2.26	-2.34	-0.95	-0.04	0.02
17	3	-0.0036	-0.44	13.06	9.75	2.88	-1.78	-0.72	-0.29	-1.96	0.76

Table 4.5: A summary of the deviations and statistical parameters for prediction of different properties

(absolute deviations for all the ASTM temperatures are in °C)

Product	Properties	Average Deviation		Maximum Deviation		R-squared value	Correlation coefficient r
		Absolute	Absolute Percentage	Absolute	Absolute Percentage		
Heavy Naphtha	IBP	3.67	3.22	8.09	7.49	0.240	0.490
	ASTM 5%	2.28	1.92	7.76	6.63	0.163	0.403
	ASTM 10%	2.55	2.09	7.00	5.83	0.063	0.252
	ASTM 90%	3.86	2.57	9.41	6.15	0.140	0.375
	ASTM 95%	4.18	2.73	8.67	5.52	0.131	0.361
	FBP	4.82	2.76	8.78	4.82	0.355	0.596
SK/ATF	Specific Gravity	0.00495	2.22	0.0142	1.75	0.671	0.820
	IBP	7.77	5.34	20.88	13.65	0.0017	-0.0414
	ASTM 5%	3.52	2.17	6.50	4.06	0.136	0.369
	ASTM 95%	3.71	1.52	10.46	3.96	0.819	0.905
	FBP	3.97	1.47	11.32	4.34	0.702	0.838
	Flash Point	1.50	3.61	3.61	9.26	0.473	0.688

Figure 4.2: Comparison of ANN predicted ASTM temperatures with laboratory measured values



4.2 Flash Point

Flash point is the temperature at which the vapor above an oil will momentarily flash or ignite in the presence of a flame. Flash point serves to indicate the temperature below which an oil can be handled without the danger of fire. It indicates the relative amount of low-boiling components present in a petroleum fraction. It also provides an idea of the range and the nature of the boiling point curve of the petroleum fraction.

The common experimental methods of determining flash point are the open cup (D92) and Penesky –Martens (D93) closed cup methods for heavy oils and Tag (D56) closed tested for lighter oils. The oil is heated at a prescribed rate in the cup (tester). An ignition spark is introduced into the tester and the temperature at each instant is recorded. The temperature, at which the vapor above the material gives the first spark, is recorded as the flash point of the oil (Nelson, 1958).

The flash point is a very important characterizing factor for the SK/ATF fraction and it has to conform to certain prescribed standards. However, just as in the case for ASTM temperatures, no online hardware sensor is available for measuring flash point and hence controlling the flash point through feedback mechanism is difficult. In the present work, an ANN model has been developed which correlates the flash point with the operating conditions and the feed TBP curve, predicted by Net-1. Net-1 along with this net connected to it in series, provides the tool for predicting flash point directly when only the operating conditions and the type of crude that is being processed are specified. Thus online prediction of flash point can be done with this model, the model being discussed in the following subsection.

4.2.1 ANN Model for Predicting Flash Point (Net-3)

Intuitively, as well as from chemical engineering knowledge, it can be expected that the flash point of SK/ATF will depend on the same input variables as its IBP and ASTM 5% temperatures. This is because all these temperatures are controlled by the lighter hydrocarbons present in the fraction and hence the same operating conditions should determine the value of each of the three - IBP, ASTM 5% and the flash point.

Inputs to the ANN model: The input variables for this net and the range of variation of these variables are presented in Table 4.2. There are 17 input variables that are fed to the net but application of principal component analysis decreased the number of input variables to 13.

Outputs from the ANN model: There is only one neuron in the output layer since there is only one output variable – the flash point.

The training, validation and the prediction sets are the same as in the other nets for SK/ATF and the outputs for the training and the validation sets are the laboratory-measured data as usual.

Network Architecture: The network architecture for this model is described in Table 4.1. The structure, without the input layer, is also shown in Figure 4.3 below.

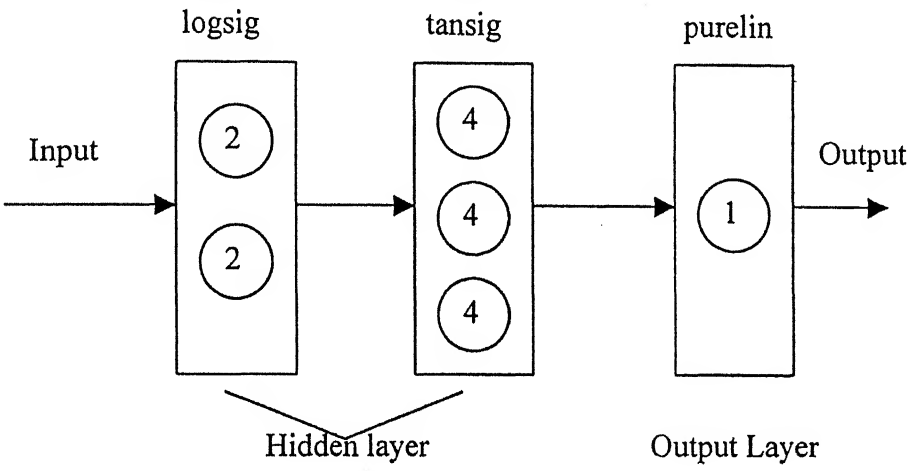


Figure 4.3:Network for predicting flash point of SK/ATF

4.2.2 Results and discussions

The values of the flash point in the prediction set range from a minimum of 38° C to a maximum of 47° C. Table 4.6 shows a comparison of ANN predicted values of the flash points for the test set with those measured in the laboratory. From this table it can be seen that the predictions are reasonably accurate within this range, the average and the maximum absolute deviations being 1.5° C and 3.6°C respectively. On the basis of its performance on the prediction set, it can be claimed that this ANN model generalizes well enough to give good predictions with input data that it has never seen before. However, the R-squared and r values are 0.473 and 0.688,

respectively (Table 4.5), which are definitely not very satisfactory. The values of these statistical parameters are expected to improve once the net is trained with more data.

Table 4.6:Flash Point of SK/ATF

Serial No.	Crude Type	Flash Point		Deviation	
		ANN output (°C)	Lab. Data (°C)	Actual (°C)	Percentage
1	1	44.52	47.0	-2.48	-5.27
2	1	43.19	41.5	1.69	4.07
3	1	43.50	42.0	1.50	3.57
4	2	38.79	41.0	-2.21	-5.39
5	2	43.27	41.5	1.77	4.27
6	2	39.24	38.5	0.74	1.92
7	2	38.92	38.0	0.92	2.42
8	3	42.09	41.0	1.09	2.66
9	1	43.19	45.0	-1.81	-4.02
10	1	43.77	43.0	0.77	1.79
11	1	39.75	41.0	-1.25	-3.05
12	1	40.13	41.0	-0.87	-2.12
13	2	42.61	39.0	3.61	9.26
14	2	44.73	44.0	0.73	1.66
15	3	43.13	40.5	2.63	6.49
16	3	43.44	43.0	0.44	1.02
17	3	.40.00	41.0	-1.00	2.44

Chapter 5

Conclusions and Recommendations

In this work, an attempt has been made to develop a package for online prediction of properties of the products from CDU using artificial neural network. Artificial Neural Networks based models have been developed to predict, online, the ASTM temperatures for Heavy Naphtha and the ASTM temperatures, the Flash Point and specific gravity for SK/ATF. The ANN models developed in this study have been tested for accuracy by applying them on previously unseen data, within the range of variation of the data for the training set. The values predicted have been compared with laboratory measured values and the maximum percentage deviations for most cases are within acceptable values except for the IBP temperature of SK/ATF fraction. The statistical parameters - R-squared and the correlation coefficient r, however, do not have values close to the desired 1, for most of the cases, but this can be somewhat explained by the presence of some stray large deviations in the predictions for each of the models. The present study establishes the potential of ANN for on-line estimation of product properties. This effort aimed to reduce the complexity of analytical equation-based calculations by capturing the model in the form of neurons interconnected by weighted links.

Further work is required before a robust and reliable model is developed. For higher accuracy in prediction, it is necessary to have, for training and validation purposes, more input data than were available to us. Also, the data, operating conditions as well as the laboratory-measured properties, are corrupted with noise as well as gross errors. This underlines the importance of re-calibration of sensors used in the industry from time to time. This also underscores the importance of developing methods for detection of gross errors in the data. Once such a method is developed, it can be used to successfully screen data for the neural net. Thus more data, free from errors, is required to tune the ANN models properly. Moreover, the data should not be available only for single crude but also during crude switches so that better generalization of the model can be achieved. Finally, once a robust package is developed for online prediction of properties, it can be used for the purpose of feedback control and on-line optimisation.

Bibliography

Basak, K.: "Opimization of a crude distillation unit", M. Tech. Dissertation, I.I.T. Kanpur, (1998).

Bhat, N.V., and McAvay, T.J.: "Use of neural networks for dynamic modeling and control of chemical process systems", Computers chem. Engng., **14**, 573 (1990).

Bhat, N. V., and McAvoy, T. J.: "Determining model structure for neural models by network stripping", Computers chem Engng., **16**, 271 (1992).

Brambilla, A. and Trivella, F.: "Estimate product quality with artificial neural network", Hydrocarbon Proc., **75.9**, 61 (1996).

Das, G.: "On-line crude TBP estimation using Artificial Neural Network", M.Tech Dissertation, I.I.T. Kanpur, (2000).

Hernandez, E., and Arkun, Y.: "Study of control-relevant properties of backpropagation neural network models of nonlinear dynamical systems", *Computers chem Engng.,* 16, 227 (1992).

Matlab Manual, Version: 5.0 " Matlab ANN Toolbox"

Murtuza, D.: "On-line backcalculation of crude feed TBP curve", M. Tech. Dissertation, I.I.T. Kanpur, (1999).

Sadhukhan, J.: "Crude charachterization and products properties estimation using Artificial Neural Networks", M.Tech Dissertation, I.I.T. Kanpur, (1997).

Satyadev, S.V.K.: "Backcalculation of true Boiling Point curve for crude oil from plant operating data", M. Tech. Dissertation, I.I.T. Kanpur,(1998).

Sharma, S.: "Estimation of phase equilibrium constants of electrolye systems using Artificial Neural Networks", M.Tech Dissertation, I.I.T. Kanpur, (1998).

Surya P. Chitra.: "Use of neural networks for problem solving", Chemical Engineering Process., p. 44, April 1993.

Watkins, R. N.: Petroleum Refinery Distillation, Gulf publishing Co. Book Division, Second edition (1981).

Appendix–1

MATLAB ANN Toolbox

The terms used for the different activation functions available in the MATLAB ANN toolbox are listed below, along with the actual mathematical functions.

Term	Full Form	Function
Compet	Competitive transfer function.	Returns output vectors with 1 where each net input vector has its maximum value, and zero elsewhere.
Hardlim	Hard limit transfer function.	hardlim(n) = 1, if n ≥ 0 0, otherwise
Hardlims	Symmetric hard limit transfer function	hardlims(n) = 1, if n ≥ 0 -1, otherwise
Logsig	Log sigmoid transfer function.	logsig(n) = 1 / (1 + exp(-n))
Poslin	Positive linear transfer function.	poslin(n) = n, if n ≥ 0 0, if n ≥ 0
Purelin	Hard limit transfer function.	purelin(n) = n
Radbas	Radial basis transfer function.	radbas(n) = exp(-n^2)
Satlin	Saturating linear transfer function.	satlin(n) = 0, if n ≤ 0 n, if 0 ≤ n ≤ 1 1, if 1 ≤ n
Satlins	Symmetric saturating linear transfer function.	satlins(n) = -1, if n ≤ -1 n, if -1 ≤ n ≤ 1 1, if 1 ≤ n
Tansig	Hyperbolic tangent sigmoid transfer function.	tansig(n) = 2/(1+exp(-2*n))-1
Tribas	Triangular basis transfer function.	tribas(n) = 1-abs(n),if -1 ≤ n ≤ 1 0, otherwise

Though the training function 'trainlm' and the performance function 'mse' have been used consistently in the present work, other training and performance functions are also available in MATLAB and are listed below.

Performance functions:

Mae	Mean absolute performance function.
Mse	Mean squared error performance function.
Msereg	Mean squared error weight/regularized performance function.
Sse	Sum squared error performance function.

Training functions: MATLAB has several training functions based on different standard optimization techniques, such as conjugate gradient and Newton methods and derive their names from these optimization techniques on which the training algorithm is based. In the following table, the different training functions, along with the optimization techniques, are listed.

Function	Optimization Technique
Trainrp	Rprop
Trainscg	Scaled Conj. Grad
Traincgf	Fletcher-Powell CG
Traincgp	Polak-Ribiere CG
Traincgb	Powell-Beale CG
Trainoss	One-Step Secant
Trainbfg	BFGS quasi-newton
Trainlm	Levenberg-Marquardt

Statistical Parameters

R-squared, coefficient of multiple determination: A statistical indicator usually applied to multiple regression analysis. It compares the accuracy of a model with the accuracy of a trivial benchmark model wherin the prediction is simply the mean of all the samples. A perfect fit between the model predictions and the actual data results in a R squared of 1, a veru good fit near 1, and a very poor fit near 0. The formula used for R-squared is:

$$R^2 = 1 - SSE / SS_{yy}$$

where $SSE = \sum (y - y')^2$, $SS_{yy} = \sum (y - y'')^2$

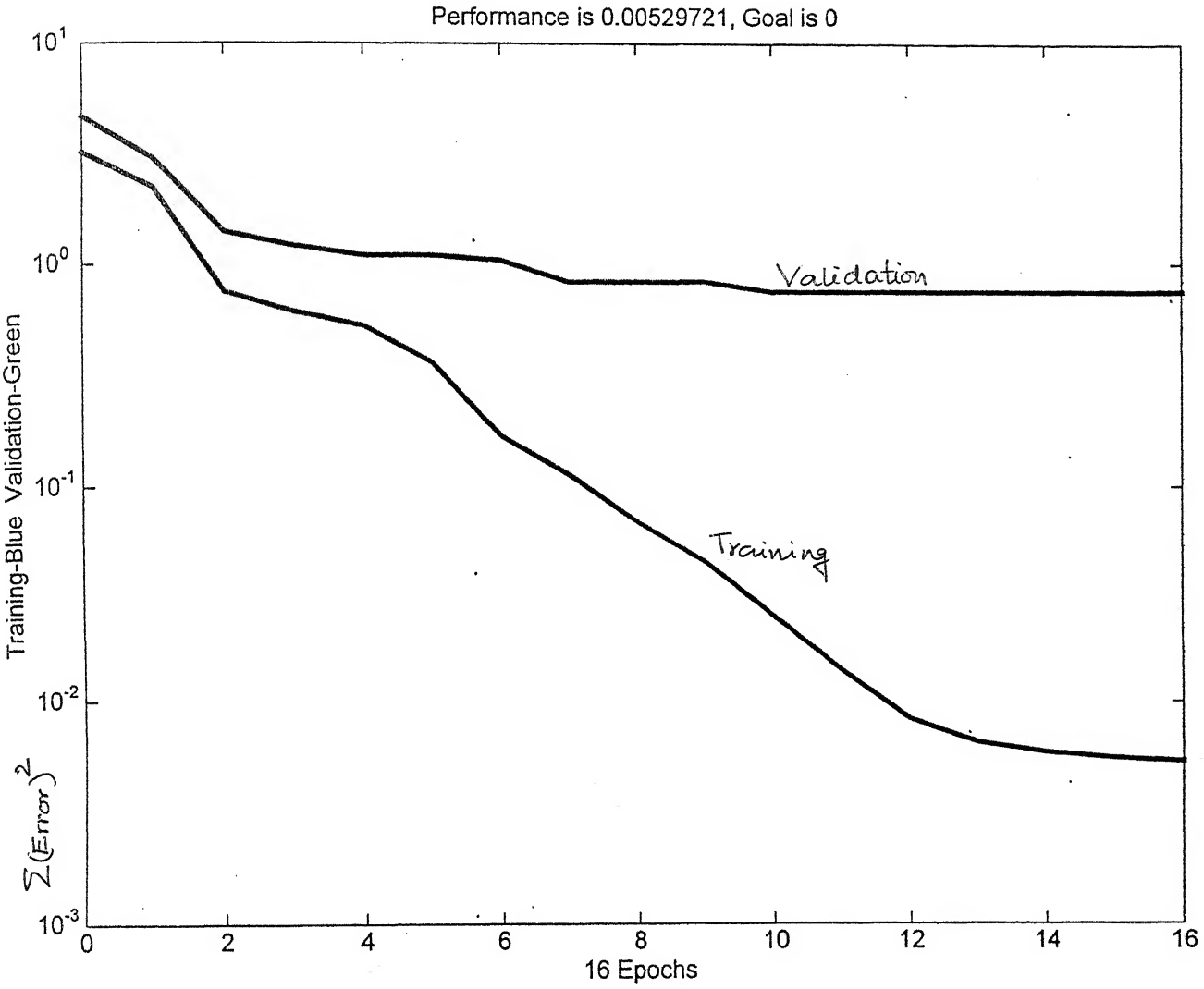
y is the actual value, y' is the predicted value, and y'' is the mean of the y values.

If the neural net predictions are worse than what one could predict by just using the sample case outputs, the R-squared value would be 1.

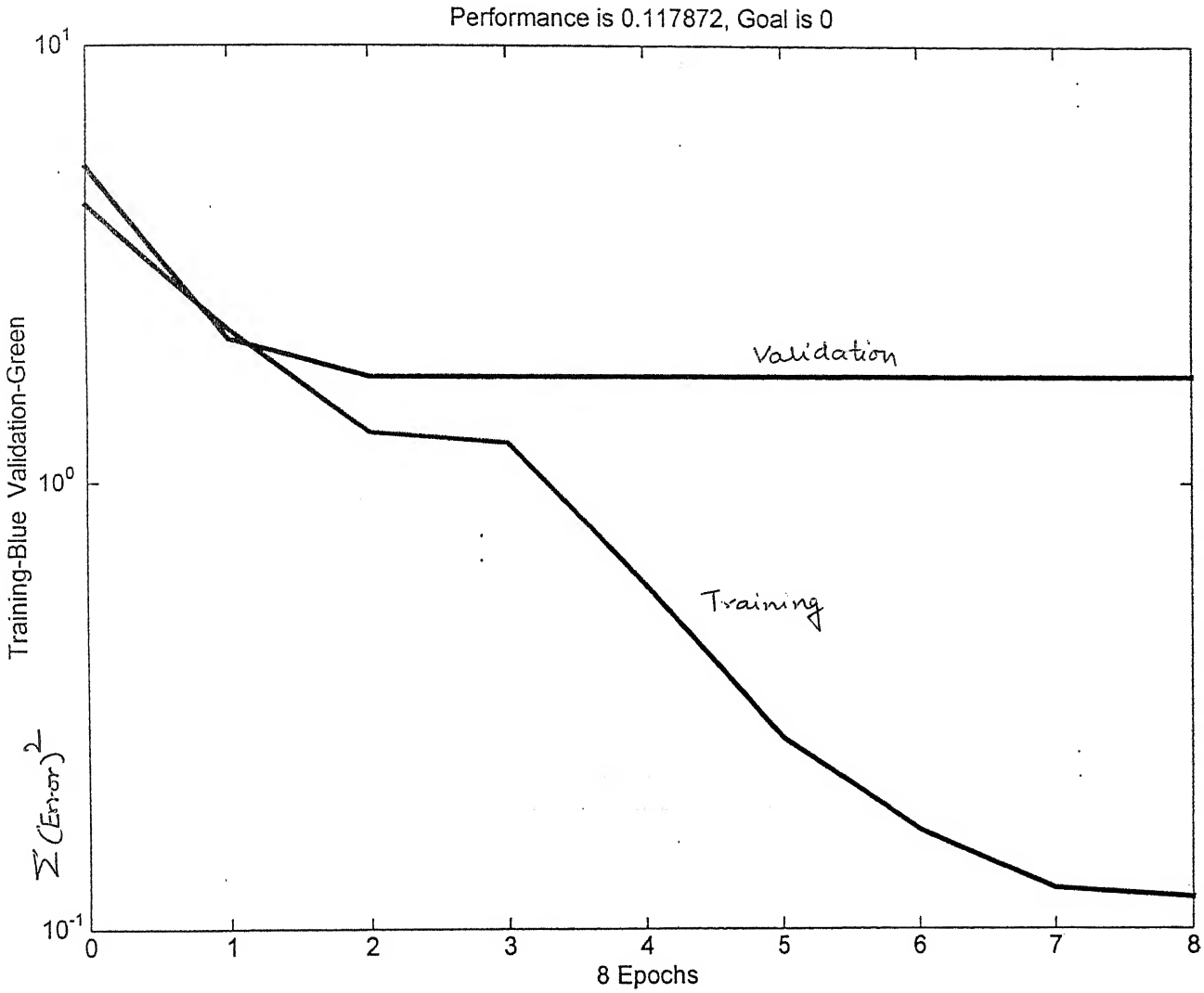
Linear Correlation coefficient, r: A statistical measure of the strength of the relationship between actual and predicted values. It can range from –1 to +1. The closer r is to +1, the stronger is the positive linear relationship, and closer it is to –1, stronger is the negative linear relationship. When r is close to 0, there is no relationship between the actual and predicted values. Basically, the correlation scatter plot is quantified by this r value.

Appendix-3

Convergence Graphs



Convergence Graph for training the net for on-line prediction of crude TBP (net-1)



Convergence Graph for training the net for on-line prediction of specific gravity, IBP and ASTM 5% of HN (net-2c)

Appendix – 4

Program Listing

The listing of all the computer programs used in the present work are available with Prof. D. N. Saraf, Department of Chemical Engineering, IIT- Kanpur.

A 131957

Date Slip

This book is to be returned on the date last stamped.